Annals of Mathematics
Vol. 39, No. 1, January, 1938

ON THE SET OF POSITIVE FUNCTIONS IN L_2

By Béla de Sz. Nagy

(Received April 14, 1937)

Introduction

Let J be a finite or infinite interval on the x-axis. Let $\alpha(x)$ be a monotone-increasing function defined on J. We shall denote by L_2 the space of all complex-valued functions, $f(x)$, defined on J, which are Lebesgue-Stieltjes-measurable with respect to $\alpha(x)$ and for which

$$\int_J |f(x)|^2 \, d\alpha(x)$$

exists; two such functions shall be considered identical if they differ at most on a null-set with respect to $\alpha(x)$.

In a previous paper[1] we have considered the following problem: Given a subset $\mathfrak{h}$ of Hilbert space $\mathfrak{H}$, what necessary and sufficient structural conditions must be satisfied by $\mathfrak{h}$ in order that a realization of $\mathfrak{H}$ by an L_2 exist which carries $\mathfrak{h}$ into the set of all characteristic functions in L_2 (i.e., into the set of all functions assuming only the values 0 or 1). A *realization* of $\mathfrak{H}$ by an L_2 is a mapping $\mathfrak{H} \leftrightarrow L_2$ which is linear and isometric (that is, if $f_1 \leftrightarrow f_1(x)$ and $f_2 \leftrightarrow f_2(x)$ are corresponding pairs, f_i in $\mathfrak{H}$, $f_i(x)$ in L_2, then $\kappa_1 f_1 + \kappa_2 f_2 \leftrightarrow \kappa_1 f_1(x) + \kappa_2 f_2(x)$ and

$$(f_1, f_2) = \int_J f_1(x) \overline{f_2(x)} \, d\alpha(x),$$

where (f_1, f_2) denotes the inner product of the elements f_1 and f_2).[2]

In the present paper we are going to give a similar structural characterization of the subset of L_2 formed by all positive functions, i.e., by all real functions $f(x) \geqq 0$.

The author is indebted to Prof. F. Riesz for having called his attention to the problem as well as for the simplifications he suggested when reading a first draft of the manuscript. In particular the form of the condition III below was suggested to the author by a paper of Prof. Riesz on the general theory of linear operations; in addition. the method of proving 5.1 (viz. differ-

---

[1] B. v. Sz. Nagy, *Über die Gesamtheit der charakteristischen Funktionen im Hilbertschen Funktionenraum*, Acta Szeged, **8** (1937), p. 166–176.

[2] $\bar{z}$ denotes the complex-conjugate of the complex number z. For the notion and fundamental properties of abstract Hilbert space see e.g. the book of M. H. Stone, *Linear transformations in Hilbert Space* (New York, 1932), Chapt. I.

entiation of certain convex functionals) is modelled on a similar argument in that paper.[3]

In the meantime the problem has also been considered in a paper by H. Freudenthal dealing with the general theory of partially-ordered moduli.[4]

Our conditions I–III below, less dependent on the general theory just mentioned, seem to be simpler than Freudenthal's and more adequate to the special needs of our problem.

1. The Conditions, Their Necessity and Independence

The purpose of this paper is to prove the following

Theorem. *Let $\mathfrak{H}$ be a Hilbert space or an n-dimensional unitary space and let $\mathfrak{P}$ be a subset of $\mathfrak{H}$. In order that there exist a realization of $\mathfrak{H}$ by an L_2 which carries $\mathfrak{P}$ into the set of all positive functions in L_2, it is necessary and sufficient that the following conditions be satisfied:*

I. $(u, v) \geqq 0$ *for every pair of elements in* $\mathfrak{P}$,

II. $(u, f) \geqq 0$, *for a fixed f in $\mathfrak{H}$ and for every u in $\mathfrak{P}$, implies $f \in \mathfrak{P}$,*

III. *if u_1, u_2, v_1, v_2 are in $\mathfrak{P}$ and if $u_1 + u_2 = v_1 + v_2$ then there exist elements $w_{11}, w_{12}, w_{21}, w_{22}$ in $\mathfrak{P}$ such that $u_i = \sum_k w_{ik}$, $v_k = \sum_i w_{ik}$ $(i, k = 1, 2)$.*

Necessity. The necessity of the conditions I–II is obvious. To prove the necessity of III let us be given four positive functions $u_1(x)$, $u_2(x)$, $v_1(x)$, $v_2(x)$ in L_2 such that $\sum_i u_i(x) = \sum_i v_i(x)$. The functions $w_{11}(x) = \inf \{u_1(x), v_1(x)\}$, $w_{12}(x) = u_1(x) - w_{11}(x)$, $w_{21}(x) = v_1(x) - w_{11}(x)$ are evidently also positive, belong to L_2 and $w_{12}(x) \cdot w_{21}(x) = 0$. As $w_{12}(x) + u_2(x) = w_{21}(x) + v_2(x)$, we have $w_{21}(x) \leqq w_{12}(x) + u_2(x)$ and therefore also $w_{21}(x) \leqq u_2(x)$. The function $w_{22}(x) = u_2(x) - w_{21}(x)$ is thus positive and it is easily seen that $u_i(x) = \sum_k w_{ik}(x)$, $v_k(x) = \sum_i w_{ik}(x)$.

Independence. In order to prove the independence of our three conditions we shall give subsets of $\mathfrak{H}$ which satisfy only two of them. $\mathfrak{H}$ itself satisfies e.g. only the conditions II–III (put $w_{ik} = \frac{1}{2}(u_i + v_k) - \frac{1}{4}(u_1 + u_2)$. On the other hand, the subset consisting of 0 alone satisfies only the conditions I and III.

It is a little more complicated to construct a subset which satisfies conditions I–II without satisfying condition III. We shall suppose that the dimension of $\mathfrak{H}$ is > 2. Let f_0 be a fixed element in $\mathfrak{H}$, $|f_0| = 1$. Let $\mathfrak{H}'$ be the (at least 2-dimensional) linear manifold consisting of all elements orthogonal to f_0. Choose a complete orthogonal set in $\mathfrak{H}'$ and denote by $\mathfrak{H}''$ the totality of all elements in $\mathfrak{H}'$ having *real* Fourier-coefficients with respect to this orthogonal set.

It is easily seen that a) (h_1, h_2) is real for h_1, h_2 in $\mathfrak{H}''$, b) if (f, h) is real for an f in $\mathfrak{H}'$ and for every h in $\mathfrak{H}''$, then f belongs to $\mathfrak{H}''$ also.

[3] F. Riesz, *A lineáris operációk általános elméletének néhány alapvető fogalomalkotásáról*, paper to appear in the Matematikai és Természettudományi Értesítö, Budapest. See also Prof. Riesz's lecture at the Bologna Congress (1928), Atti del Congresso, Vol. III, p. 143–148.

[4] H. Freudenthal, *Teilweise geordnete Moduln*, Proc. Acad. Amsterdam, **39** (1936), p. 641–651.

Let us consider now the set $\mathfrak{K}$ of all elements representable in the form

$$\kappa(f_0 + \vartheta h)$$

with $\kappa \geqq 0$, $-1 \leqq \vartheta \leqq 1$, $h \in \mathfrak{H}''$ and $|h| = 1$.

$\mathfrak{K}$ satisfies condition I:

$$(\kappa_1(f_0 + \vartheta_1 h_1),\ \kappa_2(f_0 + \vartheta_2 h_2)) = \kappa_1\kappa_2(1 + \vartheta_1\vartheta_2(h_1, h_2)) \geqq 0,$$

since $\vartheta_1\vartheta_2(h_1, h_2)$ is real and of absolute value $\leqq |h_1|\cdot|h_2| = 1$.

$\mathfrak{K}$ also satisfies condition II. For let f be an element in $\mathfrak{H}$ which has real non-negative inner products with all elements in $\mathfrak{K}$. We can always write $f = \alpha f_0 + g$ with a g in $\mathfrak{H}'$. Now as f_0 and $f_0 - h$ are in $\mathfrak{K}$ for $h \in \mathfrak{H}''$, $|h| = 1$, we obtain

$$(f, f_0) = \alpha \geqq 0 \qquad \text{and} \qquad (f, f_0 - h) = \alpha - (g, h) \geqq 0,$$

hence (g, h) is real and $(g, h) \leqq \alpha$. Each element in $\mathfrak{H}''$ being representable in the form $h \times$ *real number*, we derive from b) that g belongs to $\mathfrak{H}''$. If $g = 0$, then $f = \alpha f_0$ belongs clearly to $\mathfrak{K}$. In case $g \neq 0$ we can put $g/|g|$ for h and get

$$|g| = \left(g, \frac{g}{|g|}\right) \leqq \alpha,$$

that is

$$|g| = \vartheta\alpha, \qquad\qquad 0 < \vartheta \leqq 1.$$

We thus obtain

$$f = \alpha(f_0 + \vartheta h), \qquad\qquad (h = g/|g|)$$

and so f belongs to $\mathfrak{K}$, as we wished to prove.

We show finally that $\mathfrak{K}$ does not satisfy condition III. Choose two linearly independent elements, h_1 and h_2, in $\mathfrak{H}''$ such that $|h_1| = |h_2| = 1$ and put $u_1 = f_0 + h_1$, $u_2 = f_0 - h_1$, $v_1 = f_0 + h_2$, $v_2 = f_0 - h_2$. We then have $u_1 + u_2 = v_1 + v_2$. Let us assume that $\mathfrak{K}$ satisfies condition III, that is, that we can find four elements

$$w_{ik} = \kappa_{ik}(f_0 + \vartheta_{ik} h_{ik}) \qquad\qquad (i, k = 1, 2)$$

in $\mathfrak{K}$ so that

$$u_i = \sum_k w_{ik}, \qquad v_k = \sum_i w_{ik}.$$

We may assume without loss of generality that $w_{11} \neq 0$, thus $\kappa_{11} \neq 0$. Hence we have

$$\left.\begin{aligned} &(1) \quad u_1 = f_0 + h_1 = (\kappa_{11} + \kappa_{12})f_0 + (\kappa_{11}\vartheta_{11}h_{11} + \kappa_{12}\vartheta_{12}h_{12}),\\ &(2) \quad v_1 = f_0 + h_2 = (\kappa_{11} + \kappa_{21})f_0 + (\kappa_{11}\vartheta_{11}h_{11} + \kappa_{21}\vartheta_{21}h_{21}) \end{aligned}\right\} \text{ with } \kappa_{11} > 0.$$

From (1) we deduce

$$\kappa_{11} + \kappa_{12} = 1 \qquad \text{and} \qquad \kappa_{11}\vartheta_{11}h_{11} + \kappa_{12}\vartheta_{12}h_{12} = h_1 .$$

Applying Minkowski's inequality to the latter equation, we obtain (since $|h_{11}| = |h_{12}| = |h_1| = 1$)

$$\kappa_{11}\vartheta_{11} + \kappa_{12}\vartheta_{12} \geqq 1.$$

Since $\kappa_{11} + \kappa_{12} = 1$, $\kappa_{ik} \geqq 0$, $\vartheta_{ik} \leqq 1$, we must have $\kappa_{11}\vartheta_{11} = \kappa_{11}$ and $\kappa_{12}\vartheta_{12} = \kappa_{12}$. But the sign $=$ can occur in Minkowski's inequality only when $\kappa_{11}\vartheta_{11}h_{11} = \gamma h_1$ with $\gamma \geqq 0$. Now $\kappa_{11}\vartheta_{11} = \kappa_{11} > 0$ yields $h_{11} = h_1$.

From (2) we could deduce similarly that $h_{11} = h_2$, contrary to our initial assumption that h_1 and h_2 were linearly independent.

2. Some Consequences of Conditions I and II

In order to prove the sufficiency of our conditions we shall suppose from now on that the subset $\mathfrak{P}$ satisfies them. The letters u, v, w will always denote elements of $\mathfrak{P}$.

In the present section we shall derive some properties of $\mathfrak{P}$ using only the fact that it satisfies the conditions I–II. Two immediate consequences are the following:

2.1 *$\mathfrak{P}$ contains all linear combinations $\sum \kappa_i u_i$ with real non-negative coefficients κ_i.*

2.2. *$\mathfrak{P}$ is closed.*

Definition. If $u = v + w$, we agree to write $v \leqq u$ or $u \geqq v$.

Clearly $u_1 \leqq u_2$ and $u_2 \leqq u_3$ imply $u_1 \leqq u_3$ (use 2.1). The simultaneous relations $u \leqq v$ and $v \leqq u$ hold only for $u = v$. For these relations mean that both $v - u$ and $u - v$ belong to $\mathfrak{P}$ and so we must have, according to condition I, $(v - u, u - v) = -|v - u|^2 \geqq 0$. Thus $u = v$.

2.3. *$u_1 \leqq u_2$ and $v_1 \leqq v_2$ imply $(u_1, v_1) \leqq (u_2, v_2)$. In particular if $u_2 \perp v_2$ then $u_1 \perp v_1$ also.*

Since $u_2 - u_1$ and $v_2 - v_1$ are in $\mathfrak{P}$, we have, by condition I,

$$(u_2 - u_1, v_2) \geqq 0 \qquad \text{and} \qquad (u_1, v_2 - v_1) \geqq 0.$$

Adding these together, we obtain the desired result: $(u_2, v_2) - (u_1, v_1) \geqq 0$.

2.4. *A "monotone-decreasing" sequence:*

$$u_1 \geqq u_2 \geqq u_3 \geqq \cdots$$

is always convergent. More generally, let u_ν be a "monotone-decreasing" function of the continuously variable parameter ν, then $\lim_{\nu \to \nu_0 - 0} u_\nu$ exists.

The numerical function $|u_\nu|^2$ being $\geqq 0$ and monotone-decreasing in view of 2.3, $\lim_{\nu \to \nu_0 - 0} |u_\nu|^2$ exists. If ϵ is a given positive number, then we have

$$0 \leqq |u_\mu|^2 - |u_\nu|^2 < \epsilon \qquad \text{for} \qquad N(\epsilon) \leqq \mu < \nu < \nu_0 .$$

Using 2.3 again, we obtain

$$|u_\mu - u_\nu|^2 = |u_\mu|^2 - 2(u_\mu, u_\nu) + |u_\nu|^2 \leqq |u_\mu|^2 - 2|u_\nu|^2 + |u_\nu|^2$$
$$= |u_\mu|^2 - |u_\nu|^2 < \epsilon.$$

This proves 2.4.

NOTATION. Let $\mathfrak{R}$ denote the set of all elements r in $\mathfrak{H}$ which have real inner products with all elements of $\mathfrak{P}$.

$\mathfrak{R}$ is obviously closed and contains with two elements all their linear combinations with real coefficients.

2.5. *Every element f in $\mathfrak{H}$ is representable in the form $f = r_1 + ir_2$ with r_1, r_2 in $\mathfrak{R}$, $i = \sqrt{-1}$.*

Let f be such an element. $\mathfrak{R}$ being closed and convex (i.e., containing with r', r'' also $\frac{1}{2}(r' + r'')$), the function $\Psi(r) = |f - r|^2$ attains its minimum on $\mathfrak{R}$, at r_1 say.[5]

Let u be an arbitrarily chosen element of $\mathfrak{P}$ and let λ be a real parameter. Since $r_1 + \lambda u$ also belongs to $\mathfrak{R}$, it follows that

$$0 \leqq \Psi(r_1 + \lambda u) - \Psi(r_1) = -\lambda((f - r_1, u) + \overline{(f - r_1, u)}) + \lambda^2 |u|^2.$$

If $u \neq 0$, let us put $\dfrac{(f - r_1, u) + \overline{(f - r_1, u)}}{2 \cdot |u|^2}$ for λ. Then there follows the inequality

$$0 \leqq -\frac{((f - r_1, u) + \overline{(f - r_1, u)})^2}{4 \cdot |u|^2}.$$

[5] This can be proved along the lines of an argument first used by B. Levi in a paper on Dirichlet's problem, Rendiconti del Circolo Matematico di Palermo, **22** (1906), p. 293–360, (see particularly §7). Cf. also H. Weyl, *Die Idee der Riemannschen Fläche* (Leipzig and Berlin, 1913), p. 101. This argument has been adapted for Hilbert space (and also simplified) by F. Riesz, Acta Szeged, **7** (1934), p. 34–38.

The proof runs as follows. Let $r^{(n)}$ be a sequence of elements of $\mathfrak{R}$ for which $\Psi(r^{(n)}) \to \delta$, where δ denotes the minimum of $\Psi(r)$ on $\mathfrak{R}$. It is easily seen that the following identity holds:

$$|r^{(m)} - r^{(n)}|^2 = 2|f - r^{(m)}|^2 + 2|f - r^{(n)}|^2 - 4|f - \tfrac{1}{2}(r^{(m)} + r^{(n)})|^2.$$

$\mathfrak{R}$ being *convex*, $\frac{1}{2}(r^{(m)} + r^{(n)})$ belongs to it and therefore $|f - \frac{1}{2}(r^{(m)} + r^{(n)})|^2 = \Psi(\frac{1}{2}(r^{(m)} + r^{(n)})) \geqq \delta$. Hence

$$|r^{(m)} - r^{(n)}|^2 \leqq 2\Psi(r^{(m)}) + 2\Psi(r^{(n)}) - 4\delta \to 2\delta + 2\delta - 4\delta = 0$$

for $m, n \to \infty$. Thus $\lim r^{(n)} = r_1$ exists and, since $\mathfrak{R}$ is *closed*, it is an element of $\mathfrak{R}$. We have, using also Minkowski's inequality,

$$\delta \leqq \Psi(r_1) = |f - r_1|^2 \leqq [\,|f - r^{(n)}| + |r^{(n)} - r_1|\,]^2 \to \delta \quad \text{for } n \to \infty,$$

that is,

$$\Psi(r_1) = \delta.$$

(Observe that $\mathfrak{R}$ could mean here an arbitrary closed and convex subset in $\mathfrak{H}$.)

Hence $(f - r_1, u)$ is purely imaginary and thus $((f - r_1)/i, u)$ is real. As this remains true also for $u = 0$, we have

$$\frac{f - r_1}{i} = r_2 \in \mathfrak{R}, \qquad f = r_1 + ir_2.$$

2.6. *Every element r of $\mathfrak{R}$ is representable in a uniquely determined manner in the form $r = u_1 - u_2$, where $u_1 \perp u_2$.*

Let r be an element of $\mathfrak{R}$. The set $\mathfrak{P}$ being, by 2.1 and 2.2, closed and convex, the function $\Psi(u) = |r - u|^2$ attains on $\mathfrak{P}$ its minimum, at u_1 say.

Let u be an arbitrarily chosen element in $\mathfrak{P}$ and let κ be a positive parameter. Then $u_1 + \kappa u$ belongs also to $\mathfrak{P}$, so that

$$0 \leqq \Psi(u_1 + \kappa u) - \Psi(u_1) = -2\kappa(r - u_1, u) + \kappa^2 |u|^2.$$

Hence

$$2(r - u_1, u) \leqq \kappa |u|^2,$$

thus, if $\kappa \to 0$,

$$(r - u_1, u) \leqq 0.$$

This implies, by condition II, that $u_1 - r$ belongs to $\mathfrak{P}$, that is, $r = u_1 - u_2$.

The function

$$\Psi(\kappa u_1) = |r - \kappa u_1|^2 = |r|^2 - 2\kappa(r, u_1) + \kappa^2 |u_1|^2$$

of the positive parameter κ must take its minimal value at $\kappa = 1$, thus

$$\left.\frac{d\Psi(\kappa u_1)}{d\kappa}\right|_{\kappa=1} = -2\cdot(r, u_1) + 2|u_1|^2 = 2\cdot(u_1, u_1 - r) = 2(u_1, u_2) = 0,$$

hence $u_1 \perp u_2$.

The uniqueness of the representation $r = u_1 - u_2$ with $u_1 \perp u_2$ can be proved as follows. If

$$r = u_1 - u_2 = u_1' - u_2' \qquad \text{with} \qquad u_1 \perp u_2,\ u_1' \perp u_2',$$

then

$$u_1 - u_1' = u_2 - u_2',$$

hence (using condition I)

$$\begin{aligned} |u_1 - u_1'|^2 + |u_2 - u_2'|^2 &= 2(u_1 - u_1', u_2 - u_2') \\ &= -2(u_1', u_2) - 2(u_1, u_2') \leqq 0. \end{aligned}$$

Thus

$$u_1 = u_1' \qquad \text{and} \qquad u_2 = u_2'.$$

3. Greatest Minorants and Least Majorants

3.1. *If u_1, u_2, v_1, v_2 are such that $u_1 + u_2 = v_1 + v_2$ and $u_1 \perp v_1$, then $u_1 \leqq v_2$.*

By condition III, there exist elements w_{ik} in $\mathfrak{P}$ such that

$$u_i = \sum_k w_{ik}, \qquad v_k = \sum_i w_{ik} \qquad (i, k = 1, 2).$$

Hence

$$w_{11} \leqq u_1, \qquad w_{11} \leqq v_1.$$

From 2.3 it follows that

$$|w_{11}|^2 \leqq (u_1, v_1) = 0,$$

thus we have $w_{11} = 0$. Hence we have

$$u_1 = w_{12} \leqq w_{12} + w_{22} = v_2.$$

DEFINITION. Let u and v be two arbitrary elements in $\mathfrak{P}$. As $u - v$ belongs to $\mathfrak{R}$, we may represent it, according to 2.6, in the form $u' - v'$ with $u' \perp v'$. Denote $u - u'$ (which is equal to $v - v'$) by $\inf\{u, v\}$ and $u + v - \inf\{u, v\}$ (or, what is the same, $u' + v' + \inf\{u, v\}$) by $\sup\{u, v\}$. The uniqueness of the above representation implies that $\inf\{u, v\}$ and $\sup\{u, v\}$ are uniquely determined by u and v, without respect to their order.

3.2. *The elements* $\inf\{u, v\}$ *and* $\sup\{u, v\}$ *belong to* $\mathfrak{P}$ *and enjoy properties similar to those of the greatest minorants and least majorants (or envelopes) of two positive functions, that is*

1. $\inf\{u, v\} + \sup\{u, v\} = u + v$,
2. $\sup\{u, v\} \geqq \begin{Bmatrix} u \\ v \end{Bmatrix} \geqq \inf\{u, v\}$,
3. *if* $w \geqq u$ *and* $w \geqq v$ *simultaneously then* $w \geqq \sup\{u, v\}$,
4. *if* $w \leqq u$ *and* $w \leqq v$ *simultaneously then* $w \leqq \inf\{u, v\}$,
5. $v - \inf\{u, v\} = \sup\{u, v\} - u \perp \sup\{u, v\} - v = u - \inf\{u, v\}$.

Write $u - v$ in the form $u' - v'$ with $u' \perp v'$. Applying 3.1 to the equation $u + v' = u' + v$, we obtain the result that $v' \leqq v$. Now $\inf\{u, v\}$ being equal to $v - v'$ and $\sup\{u, v\}$ being equal to $u' + v' + \inf\{u, v\}$, both belong to $\mathfrak{P}$.

Clearly 1 and 5 are immediate consequences of the definition.

Since $u - u' = v - v'$ belongs to $\mathfrak{P}$, it follows that $\inf\{u, v\} = u - u' \leqq u$. Similarly $\inf\{u, v\} \leqq v$. On the other hand,

$$\sup\{u, v\} = u + [v - \inf\{u, v\}] \geqq u$$

and similarly $\sup\{u, v\} \geqq v$; thus 2 is also verified.

Assume that $w \geqq u$, $w \geqq v$ simultaneously. We have then also

$$w - \inf\{u, v\} \geqq \begin{cases} u - \inf\{u, v\} = u' \\ v - \inf\{u, v\} = v'. \end{cases}$$

This means that there exist two elements, u'' and v'' say, such that

$$w - \inf\{u, v\} = u' + u'' = v' + v''.$$

As $u' \perp v'$, we derive from 3.1 that $u' \leqq v''$. Thus

$$w - \inf\{u, v\} = v' + v'' \geqq v' + u',$$

hence

$$w \geqq u' + v' + \inf\{u, v\} = \sup\{u, v\}.$$

So we have proved 3 also.

Assume now that $w \leqq u$ and $w \leqq v$ hold simultaneously. Then $u \leqq u + (v - w)$ and $v \leqq v + (u - w)$ and this implies, by 3, that

$$u + v - w \geqq \sup \{u, v\} = u + v - \inf \{u, v\}.$$

Hence

$$w \leqq \inf \{u, v\},$$

as we asserted in **4**.

3.3. *If* $w \geqq u$ *and* $w \geqq v$ *hold simultaneously then*

$$w - \sup \{w - u, w - v\} = \inf \{u, v\}.$$

First, we have $w \geqq w - u$, $w \geqq w - v$ and therefore, by 3.2.3, also $w \geqq \sup \{w - u, w - v\}$. Furthermore, as

$$\sup \{w - u, w - v\} \geqq \begin{cases} w - u \\ w - v, \end{cases}$$

we obtain that

$$w - \sup \{w - u, w - v\} \leqq \begin{cases} u \\ v \end{cases}.$$

Thus, by 3.2.4,

$$w - \sup \{w - u, w - v\} \leqq \inf \{u, v\}. \tag{3}$$

Next, since

$$w - \inf \{u, v\} \geqq \begin{cases} w - u \\ w - v \end{cases}$$

(use 3.2.2), it follows from 3.2.3 again that

$$w - \inf \{u, v\} \geqq \sup \{w - u, w - v\},$$

hence

$$w - \sup \{w - u, w - v\} \geqq \inf \{u, v\}.$$

This, together with (3), implies indeed:

$$w - \sup \{w - u, w - v\} = \inf \{u, v\}.$$

4. Definition and Some Properties of the Set $\mathfrak{h}$

We have introduced in the cited paper (see footnote 1) the notation $f_1 \prec f_2$ for two elements in $\mathfrak{H}$ connected by the relation

$$f_1 \perp f_2 - f_1 .$$

Let now $u^{(1)}, u^{(2)}, u^{(3)}, \cdots, u^{(k)}, \cdots$ be a sequence of elements everywhere dense in $\mathfrak{P}$ and let us put

$$h^* = \sum_{k=1}^{\infty} \frac{u^{(k)}}{2^k \cdot | u^{(k)} |}.$$

From 2.1 and 2.2 we deduce that h^* belongs to $\mathfrak{P}$.

NOTATION. Let $\mathfrak{h}$ be the set of all elements u of $\mathfrak{P}$ such that

$$u \leqq h^* \quad \text{and} \quad u \prec h^*.$$

Plainly h^* itself belongs to $\mathfrak{h}$. If h belongs to it, then $h^* - h$ does also. Elements denoted by the letter h will be always supposed belonging by $\mathfrak{h}$.

4.1. *If the difference of h_2 and h_1, that is $h_2 - h_1$, belongs also to $\mathfrak{h}$, then $h_1 \prec h_2$.*

From the assumed relation

$$h_2 - h_1 \perp h^* - (h_2 - h_1) = (h^* - h_2) + h_1,$$

it follows, since $h_1 \leqq (h^* - h_2) + h_1$, that indeed

$$h_2 - h_1 \perp h_1$$

(use 2.3).

4.2. *Both* $\inf \{h_1, h_2\}$ *and* $\sup \{h_1, h_2\}$ *belong to* $\mathfrak{h}$ *and* $\inf \{h_1, h_2\} \prec h_1$, $\inf \{h_1, h_2\} \prec h_2$.

It follows from 3.2.3 that $h^* \geqq \sup \{h_1, h_2\}$. Moreover, we have

$$h^* - \sup \{h_1, h_2\} \leqq h^* - h_1 \quad \text{and} \quad h^* - \sup \{h_1, h_2\} \leqq h^* - h_2.$$

But

$$h_1 \perp h^* - h_1 \quad \text{and} \quad h_2 \perp h^* - h_2$$

imply, then, according to 2.3, that

$$h_1 \perp h^* - \sup \{h_1, h_2\} \quad \text{and} \quad h_2 \perp h^* - \sup \{h_1, h_2\}.$$

Hence

$$h_1 + h_2 \perp h^* - \sup \{h_1, h_2\}.$$

But as $\sup \{h_1, h_2\} = h_1 + h_2 - \inf \{h_1, h_2\} \leqq h_1 + h_2$, we obtain, using 2.3 again:

$$\sup \{h_1, h_2\} \perp h^* - \sup \{h_1, h_2\},$$

that is

$$\sup \{h_1, h_2\} \prec h^*.$$

We have thus proved that $\sup \{h_1, h_2\}$ belongs to $\mathfrak{h}$.

Applying this result to $h^* - h_1$ and $h^* - h_2$ instead of h_1 and h_2, we obtain that $\sup \{h^* - h_1, h^* - h_2\}$ and therefore

$$h^* - \sup \{h^* - h_1, h^* - h_2\}$$

also belong to $\mathfrak{h}$. The latter being equal to $\inf \{h_1, h_2\}$ in view of 3.3, $\inf \{h_1, h_2\}$ also belongs to $\mathfrak{h}$.

The relations

$$\inf \{h_1, h_2\} \perp h^* - \inf \{h_1, h_2\} \quad \text{and} \quad h^* - \inf \{h_1, h_2\} \geqq h_1 - \inf \{h_1, h_2\}$$

imply, again by 2.3, that

$$\inf \{h_1, h_2\} \prec h_1 .$$

Changing the rôle of h_1 and h_2, we get

$$\inf \{h_1, h_2\} \prec h_2 .$$

So we have verified all statements of 4.2.

4.3. *If $h_1 \prec h_2$, then $h_2 - h_1$ also belongs to $\mathfrak{h}$.*

We have seen just now, that $h_3 = \inf \{h_1, h_2\}$ belongs to $\mathfrak{h}$ and that

$$h_3 \perp h_1 - h_3 , \tag{4}$$

$$h_3 \perp h_2 - h_3 . \tag{5}$$

On the other hand, 3.2.5 gives

$$h_1 - h_3 \perp h_2 - h_3 .$$

This and (5) imply

$$h_1 = (h_1 - h_3) + h_3 \perp h_2 - h_3 .$$

This and the assumed relation: $h_1 \prec h_2$, that is,

$$h_1 \perp h_2 - h_1$$

imply further

$$h_1 \perp (h_2 - h_3) - (h_2 - h_1) = h_1 - h_3 .$$

This and (4) yield finally

$$h_1 - h_3 \perp h_1 - h_3 ,$$

that is to say,

$$h_1 = h_3 = \inf \{h_1, h_2\}.$$

It follows that $h_2 - h_1$ belongs to $\mathfrak{P}$. We have also

$$h_2 - h_1 \leqq h_2 \leqq h^*.$$

In order to complete the proof of 4.3, we need still to show that

$$h_2 - h_1 \prec h^*.$$

Now, we have by supposition $h_1 \prec h_2$ and $h_2 \prec h^*$, that is,

$$h_2 - h_1 \perp h_1 \quad \text{and} \quad h_2 \perp h^* - h_2 .$$

We may replace h_2 on the left-hand side of the second relation by $h_2 - h_1$, since $h_2 - h_1 \leqq h_2$ (use 2.3). The left-hand sides thus becoming equal, we can add the right-hand sides together. So we obtain the desired result:

$$h_2 - h_1 \perp h_1 + (h^* - h_2) = h^* - (h_2 - h_1).$$

4.4. *The set $\mathfrak{h}$ is closed.*

This follows easily from the closure of the set $\mathfrak{P}$ and from the continuity of the inner product.

5. The Completeness of $\mathfrak{h}$ in $\mathfrak{H}$

Let u be an arbitrarily chosen element in $\mathfrak{P}$ and let λ be a real non-negative parameter. Then

$$s_\lambda = \sup \{\lambda h^*, u\}$$

is a monotone-increasing function of the parameter, while

$$s_\lambda - \lambda h^*,$$

being equal to $u - \inf \{\lambda h^*, u\}$, is monotone-decreasing.

For $0 \leqq \kappa \leqq 1$ we obtain the inequalities:

$$(1 - \kappa)s_\lambda + \kappa s_\mu \geqq \begin{cases} (1 - \kappa)\lambda h^* + \kappa\mu h^* = [(1 - \kappa)\lambda + \kappa\mu]h^* \\ (1 - \kappa)u + \kappa u = u \end{cases}.$$

Hence, by 3.2.3,

$$(1 - \kappa)s_\lambda + \kappa s_\mu \geqq s_{(1-\kappa)\lambda+\kappa\mu},$$

s_λ is thus a convex function of λ.

The monotoneity of s_λ implies that $\dfrac{s_\mu - s_\lambda}{\mu - \lambda}$ belongs to $\mathfrak{P}$. The convexity of s_λ implies further that this quotient decreases monotonely when μ approaches from the right to the fixed λ. So we may apply 2.4, which assures us that

$$h_\lambda = \lim_{\mu\to\lambda+0} \frac{s_\mu - s_\lambda}{\mu - \lambda}$$

exists (and belongs to $\mathfrak{P}$ because of the closure of $\mathfrak{P}$). We shall show that it belongs also to $\mathfrak{h}$ and so the notation with the letter h will be justified.

If $\nu > 0$ then $s_{\lambda+\nu} - s_\lambda$ belongs to $\mathfrak{P}$ and plainly

$$s_{\lambda+\nu} - u \geqq s_{\lambda+\nu} - s_\lambda . \tag{6}$$

The function $s_\lambda - \lambda h^*$ being monotone-decreasing, we have also

$$\begin{aligned} s_{\lambda+\nu} - (\lambda + \nu)h^* &\geqq (s_{\lambda+\nu} - (\lambda + \nu)h^*) - (s_{\lambda+2\nu} - (\lambda + 2\nu)h^*) \\ &= s_{\lambda+\nu} - s_{\lambda+2\nu} + \nu h^*. \end{aligned} \tag{7}$$

The right-hand sides of (6) and (7) are orthogonal in view of 3.2.5. Applying 2.3, we have thus

$$s_{\lambda+\nu} - s_\lambda \perp \nu h^* - (s_{\lambda+2\nu} - s_{\lambda+\nu})$$

and therefore

$$\frac{s_{\lambda+\nu} - s_\lambda}{\nu} \perp \frac{\nu h^* - (s_{\lambda+2\nu} - s_{\lambda+\nu})}{\nu} = h^* - 2\frac{s_{\lambda+2\nu} - s_\lambda}{2\nu} + \frac{s_{\lambda+\nu} - s_\lambda}{\nu}.$$

For $\nu \to 0$ this becomes

$$h_\lambda \perp h^* - 2h_\lambda + h_\lambda = h^* - h_\lambda .$$

We have thus the following double result: $h^* - h_\lambda$ belongs to $\mathfrak{P}$ and $h_\lambda \prec h^*$. Thus h_λ belongs to $\mathfrak{h}$ indeed.

From the convexity of s_λ it follows easily that h_λ is monotone-increasing and, moreover, that

$$(\mu - \lambda)h_\lambda \leqq s_\mu - s_\lambda \leqq (\mu - \lambda)h_\mu \qquad \text{for } \mu \geqq \lambda. \tag{8}$$

Suppose now that u is one of the elements of the sequence $u^{(k)}$ used in the definition of h^*, then there exists plainly a number Λ for which $u \leqq \Lambda h^*$ (if $u = u^{(k)}$, then $\Lambda = 2^k \,|\, u^{(k)} \,|$).

We have then obviously

$$s_0 = \sup \{0, u\} = u,$$

$$\left.\begin{array}{l} s_\lambda = \sup \{\lambda h^*, u\} = \lambda h^* \\ h_\lambda = h^* \end{array}\right\} \text{for } \lambda \geqq \Lambda.$$

Let ϵ be a given positive number and let n be an integer such that

$$n \geqq (\Lambda/\epsilon) \,|\, h^* - h_0 \,|.$$

Put

$$h_{(m/n)\Lambda} = h^{(m)}, \qquad s_{(m/n)\Lambda} = s^{(m)} \qquad \text{for } m = 0, 1, \cdots, n.$$

It then follows from (8) that

$$\frac{\Lambda}{n} h^{(m)} \leqq s^{(m+1)} - s^{(m)} \leqq \frac{\Lambda}{n} h^{(m+1)}, \quad m = 0, 1, \cdots, n-1,$$

and so

$$\frac{\Lambda}{n} \sum_{m=0}^{n-1} h^{(m)} \leqq \sum_{m=0}^{n-1} (s^{(m+1)} - s^{(m)}) = \Lambda h^* - u \leqq \frac{\Lambda}{n} \sum_{m=0}^{n-1} h^{(m+1)}.$$

Thus

$$\frac{\Lambda}{n} \sum_{m=0}^{n-1} h^{(m+1)} - (\Lambda h^* - u) \leqq \frac{\Lambda}{n} \sum_{m=0}^{n-1} h^{(m+1)} - \frac{\Lambda}{n} \sum_{m=0}^{n-1} h^{(m)}.$$

The left-hand side is equal to

$$u - \Lambda(h^* - \frac{1}{n} \sum_{m=0}^{n-1} h^{(m+1)}) = u - \frac{\Lambda}{n} \sum_{m=0}^{n-1} (h^* - h^{(m+1)}),$$

while the right-hand side is equal to

$$\frac{\Lambda}{n} \sum_{m=0}^{n-1} (h^{(m+1)} - h^{(m)}) = \frac{\Lambda}{n} (h^{(n)} - h^{(0)}) = \frac{\Lambda}{n} (h^* - h_0).$$

We have therefore

$$u - \frac{\Lambda}{n} \sum_{m=0}^{n-1} (h^* - h^{(m+1)}) \leqq \frac{\Lambda}{n} (h^* - h_0).$$

Hence, by 2.3,

$$\left| u - \frac{\Lambda}{n} \sum_{m=0}^{n-1} (h^* - h^{(m+1)}) \right|^2 \leqq \frac{\Lambda^2}{n^2} | h^* - h_0 |^2 \leqq \epsilon^2.$$

Since $h^{(m+1)}$ and therefore $h^* - h^{(m+1)}$ belong to $\mathfrak{h}$, we have shown that every element of the sequence $u^{(1)}, u^{(2)}, u^{(3)}, \cdots$ may be approximated, as closely as desired, by linear combinations of h's with positive coefficients.

The sequence $u^{(1)}, u^{(2)}, u^{(3)}, \cdots$ being everywhere dense in $\mathfrak{P}$, we have more generally:

5.1. *Every element in $\mathfrak{P}$ may be approximated, as closely as desired, by linear combinations of elements of $\mathfrak{h}$ with positive coefficients.*

This result and 2.5, 2.6 imply also the following:

5.2. *$\mathfrak{h}$ is a complete subset of $\mathfrak{H}$, i.e., the linear combinations formed of elements of $\mathfrak{h}$ with complex coefficients are everywhere dense in $\mathfrak{H}$.*

6. Realization by an L_2

The properties of $\mathfrak{h}$ established in 4.1, 4.2 (see also 3.2.1), 4.3, 4.4 and 5.2 imply, by a previous theorem of ours,[6] that there exists a realization of $\mathfrak{H}$ by an L_2 which carries $\mathfrak{h}$ into the set of all characteristic functions in L_2.

We shall show that this realization carries $\mathfrak{P}$ into the set of all positive functions in L_2.

Let u be an arbitrary element of $\mathfrak{P}$. Suppose this is carried into the function $f(x)$ in L_2. Since $(u, h) \geqq 0$ for every h in $\mathfrak{h}$, it follows from the isometrical property of the realization that $f(x)$ also has real non-negative product-integrals with every characteristic function belonging to L_2. Hence $f(x)$ must be $\geqq 0$.

Conversely, let f be such an element of $\mathfrak{H}$ which is carried into a positive function $u(x)$ in L_2. Since the product of $u(x)$ by an arbitrary characteristic function in L_2 has plainly a real non-negative integral, it follows again from the isometrical property of the realization that $(f, h) \geqq 0$ for every h in $\mathfrak{h}$. We have then, in view of 5.1, also $(f, u) \geqq 0$ for every u in $\mathfrak{P}$, thus, by condition II, f belongs to $\mathfrak{P}$ itself.

Thus we have also proved the part of our theorem asserting the sufficiency of conditions I–III.

Szeged, Hungary.

[6] Cf. the cited paper. This theorem establishes conditions for a subset $\mathfrak{C}$ of $\mathfrak{H}$ in order that there exist a realization of $\mathfrak{H}$ by an L_2 which carries $\mathfrak{C}$ into the set of all characteristic functions in L_2. They are the following: a) $\mathfrak{C}$ is a complete subset of $\mathfrak{H}$; b) the difference $f - g$ of two elements of $\mathfrak{C}$ belongs also to $\mathfrak{C}$ if and only if $g \prec f$; c) if f_1, f_2 are in $\mathfrak{C}$ then there are other two elements g_1, g_2 in $\mathfrak{C}$ such that $f_1 + f_2 = g_1 + g_2$ and $g_1 \prec f_1$, $g_1 \prec f_2$; d) $\mathfrak{C}$ is closed.

Annals of Mathematics
Vol. 39, No. 1, January, 1938

A THEOREM ON ANALYTIC CONTINUATION OF FUNCTIONS IN SEVERAL VARIABLES

By S. Bochner

(Received July 23, 1937)

Introduction. As in a previous paper[1] we shall consider the k-dimensional Euclidean space of the real points $x = (x_1, \cdots, x_k)$ and its complex extension consisting of the complex points $z = (z_1, \cdots, z_k)$; $z_\kappa = x_\kappa + iy_\kappa$. Any point set S of the real space is the basis of a *tube* $T = T_S$ of the complex space. The latter is the point set of the complex space consisting of all k-dimensional planes

$$x_\kappa = x_\kappa^0; \qquad -\infty < y_\kappa < \infty, \quad \kappa = 1, \cdots, k,$$

where $(x_1^0, \cdots, x_\kappa^0)$ is an arbitrary point of S. A tube is an (open) domain if and only if its basis is a domain. We shall say that a tube T' lies *within* a tube T if the closure of T' is part of the interior of T. The convex closure (in the usual sense) of the tube T will be denoted by $\tilde{T}$. Obviously $\tilde{T}$ is again a tube and its base $\tilde{S}$ is the convex closure of S.

In the paper loc. cit. it was shown that if a function $f(z) = f(z_1, \cdots, z_k)$ is analytic and bounded within T it also exists and is bounded within $\tilde{T}$. In the present paper we shall establish the theorem omitting the property of boundedness.

Theorem. *Any function which is analytic in a tube T is analytic in its convex closure $\tilde{T}$.*

On the other hand, any convex tube $\tilde{T}$ is the natural domain of analyticity for some function. In fact, there exists functions of this kind which have the period $2\pi i$ in each variable z_κ. Combining this fact with our theorem we are led to the following statement: *the envelope of regularity* (*Regularitätshülle*)[2] *of T is $\tilde{T}$.*

From the viewpoint of the theory of analytic functions of real variables our theorem may be formulated as follows. A function $f(x_1, \cdots, x_k)$ which is analytic over a domain S has, by definition, an extension

$$f(x_1 + iy_1, \cdots, x_k + iy_k)$$

into a complex neighborhood of S. If, now, the extended neighborhood is the whole tube T_S, then $f(x_1, \cdots, x_k)$ has an analytic continuation into the convex closure $\tilde{S}$ of S (and its tube $\tilde{T}$).

---

[1] *Bounded analytic functions in several variables and multiple Laplace integrals.* American Jour. of Math., vol. 59 (1937), pp. 732–738.

[2] H. Behnke und P. Thullen, *Theorie der Funktionen mehrerer komplexer Veränderlichen*, Ergebnisse der Mathematik, vol. 3 (1934), p. 70.

Our method of proof seems to be new. We shall approximate to the infinite tube T by finite regions with elliptic cross-sections and expand $f(z)$ in multiple series of Legendre polynomials. In this way we shall prove the analyticity of $f(z)$ in a tube T^* containing T and, finally, we shall show that T^* is identical with $\tilde{T}$.

Expansions in multiple Legendre series. Denoting by $C(r_\kappa) \equiv C(r_1, \cdots, r_k)$ the circular polycylinder

$$(1) \qquad |z_\kappa| < r_\kappa \qquad \kappa = 1, \cdots, k,$$

if $f(z)$ is analytic in a neighborhood of the origin containing the polycylinders $C(r'_\kappa)$, $C(r''_\kappa)$, it also is analytic in each polycylinder $C(r_\kappa)$, where

$$(2) \qquad \log r_\kappa = \alpha \log r'_\kappa + (1 - \alpha) \log r''_\kappa, \qquad 0 \leqq \alpha \leqq 1.$$

This statement on the analyticity of $f(z)$ follows immediately from the corresponding statement on the absolute convergence of its power series,[3] ,

$$(3) \qquad \sum a_{n_1 \cdots n_k} z_1^{n_1} \cdots z_k^{n_k}.$$

Replacing (1) by

$$(4) \qquad |z_\kappa + (z_\kappa^2 - 1)^{\frac{1}{2}}| < r_\kappa \qquad \kappa = 1, \cdots, k$$

we obtain a polycylinder $E(r_\kappa)$ whose components are ellipses with foci at the points $z_\kappa = +1, -1$, the quantity r_κ being the sum of the two semi-axes of the κ-th ellipse.

LEMMA. *If $f(z)$ is analytic in a neighborhood of the rectangle*

$$-1 \leqq x_\kappa \leqq +1, \qquad \kappa = 1, \cdots, k$$

containing the polycylinders $E(r'_\kappa)$, $E(r''_\kappa)$, it also is analytic in each polycylinder $E(r_\kappa)$ whose radii belong to the family (2).

PROOF. We consider the Legendre polynomials $P_n(z)$ and the Legendre functions of the second kind $Q_n(\zeta)$, and we write

$$w = z + (z^2 - 1)^{\frac{1}{2}}, \qquad \omega = \zeta + (\zeta^2 - 1)^{\frac{1}{2}}.$$

It is known that[4] for fixed $\delta > 0$, $M > 0$,

$$\frac{1}{\zeta - z} = \sum_{n=0}^{\infty} (2n + 1) P_n(z) Q_n(\zeta),$$

uniformly in

$$|w| \leqq |\omega| - \delta \leqq M,$$

[3] Behnke-Thullen, l. c., p. 37, pp. 74–75.

[4] E. W. Hobson, *Spherical and ellipsoidal harmonics*, Cambridge (1931), pp. 60–62.

and that

$$| P_n(z) | \leqq | w |^n \tag{5}$$

$$| Q_n(\zeta) | \leqq \left(\frac{\pi}{n}\right)^{\frac{1}{2}}\left(1 - \frac{1}{| \omega |^2}\right)^{-\frac{1}{2}} \cdot \frac{1}{| \omega |^{n+1}}$$

(6)

$$Q_0(\zeta) = \tfrac{1}{2} \log \frac{\zeta + 1}{\zeta - 1}.$$

If $f(z)$ is analytic in the closure of $E(\rho_\kappa)$, we can apply Cauchy's formula

$$f(z_1, \cdots, z_k) = \frac{1}{(2\pi i)^k} \int \cdots \int \frac{f(\zeta_1, \cdots, \zeta_k)\, d\zeta_1 \cdots d\zeta_k}{(\zeta_1 - z_1) \cdots (\zeta_k - z_k)}$$

in which the integrations extend over the boundaries of the ellipses. The substitution

$$\frac{1}{\zeta_\kappa - z_\kappa} = \sum_{n=0}^{\infty} (2n + 1) P_n(z_\kappa) Q_n(\zeta_\kappa)$$

leads to an expansion

$$f(z) = \sum a_{n_1 \cdots n_k} P_{n_1}(z_1) \cdots P_{n_k}(z_k) \tag{7}$$

which is absolutely and uniformly convergent in every closed subset of $E(\rho_\kappa)$; and, on account of (6), the coefficients satisfy an inequality

$$| a_{n_1 \cdots n_k} | \leqq K \cdot \rho_1^{-n_1} \cdots \rho_k^{-n_k}$$

in which the constant K is independent of the indices $n_1, \cdots, n_k$.

As in the case of power series (3), an expansion of the form (7) which converges absolutely and uniformly in an elliptic polycylinder is unique. Hence, under the assumptions of our lemma, there exists an expansion (7) such that, for fixed $\epsilon > 0$,

$$| a_{n_1 \cdots n_k} | \leqq K \cdot (r_1' - \epsilon)^{-n_1} \cdots (r_k' - \epsilon)^{-n_k}$$

$$| a_{n_1 \cdots n_k} | \leqq K \cdot (r_1'' - \epsilon)^{-n_1} \cdots (r_k'' - \epsilon)^{-n_k}.$$

If a system of coefficients satisfies these two inequalities simultaneously, the series

$$\sum | a_{n_1 \cdots n_k} | r_1^{n_1} \cdots r_k^{n_k}$$

converges for any system of radii for which

$$\log r_\kappa < \alpha \log (r_\kappa' - \epsilon) + (1 - \alpha) \log (r_\kappa'' - \epsilon).$$[5]

[5] H. Tietze, *Über den Bereich absoluter Konvergenz von Potenzreihen mehrerer Veränderlichen*, Math. Annalen, vol. 99 (1928), pp. 181–182.

Letting $\epsilon \to 0$, the assertion of our lemma becomes an immediate consequence of property (5).

In the following section we shall replace the polycylinder $E(r_\kappa)$ by the polycylinder $E_l(r_\kappa)$ whose components are ellipses with foci at the points $z_\kappa = +il$, $-il$, the quantity r_κ being again the sum of the two semi-axes of the k-th ellipse. The quantity l is a fixed positive number. It is easily seen that the lemma remains valid.

A special case. We shall first prove our theorem for a tube whose basis S is the sum of two (k-dimensional) rectangles

$$(8) \qquad | x_\kappa | < a'_\kappa \qquad \kappa = 1, \cdots, k$$

$$(9) \qquad | x_\kappa | < a''_\kappa \qquad \kappa = 1, \cdots, k.$$

Since a linear transformation of the x-coordinates can be interpreted as a linear transformation of the z-coordinates and carries tubes into tubes we are actually dealing with the case of a basis S being the sum of two rectangles which are concentric and coaxial.

The tubes (8), (9) contain the polycylinders $E_l(r'_\kappa)$, $E_l(r''_\kappa)$ for which

$$r'_\kappa = a'_\kappa + (a'^2_\kappa + l^2)^{\frac{1}{2}}, \qquad r''_\kappa = a''_\kappa + (a''^2_\kappa + l^2)^{\frac{1}{2}}.$$

By our lemma $f(z)$ is analytic also in the polycylinders $E_l(r_\kappa)$ whose small semi-axes a_κ satisfy the relations

$$\log (a_\kappa + (a^2_\kappa + l^2)^{\frac{1}{2}} = \alpha \log (a'_\kappa + (a'^2_\kappa + l^2)^{\frac{1}{2}} + (1 - \alpha) \log (a''_\kappa + (a''^2_\kappa + l^2)^{\frac{1}{2}}$$

or, what is the same,

$$(10) \qquad \operatorname{arcsinh} \frac{a_\kappa}{l} = \alpha \operatorname{arcsinh} \frac{a'_\kappa}{l} + (1 - \alpha) \operatorname{arcsinh} \frac{a''_\kappa}{l}, \qquad 0 \leqq \alpha \leqq 1.$$

Letting $l \to \infty$, relation (10) goes over into

$$a_\kappa = \alpha a'_\kappa + (1 - \alpha) a''_\kappa,$$

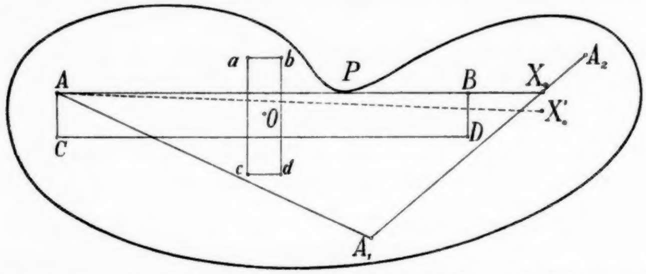
and the polycylinder $E_l(r_\kappa)$ converges to the corresponding tube

$$| x_\kappa | < a_\kappa .$$

The totality of these tubes for all values α is the convex closure of the tubes (8), (9) and our theorem is proved in this case.

For general tubes we easily conclude that if $f(z)$ is analytic in the tube T it also exists and is analytic in the tube T^* whose basis S^* is the smallest (connected) neighborhood of S having the following property: if a subdomain σ of S^* is the sum of two rectangles which are concentric and coaxial, then the convex closure $\tilde{\sigma}$ of σ is likewise a subdomain of S.

A characterization of convex domains. Finally we have to show that the domain S^* is convex in the usual sense. Any two points A, A_n of S can be connected by a (rectilinear) polygon $AA_1A_2 \cdots A_n$ and we have to prove that the polygon can be replaced by the segment AA_n. Applying induction it will be sufficient to prove that the segment AA_2 lies in S. We shall assume the contrary and derive a contradiction.[6] In order to simplify the argument we shall first treat the two-dimensional case $k = 2$. Let X be a point of A_1A_2 moving from A_1 to A_2 and let X_0 be the first value of X for which the segment AX_0 is not contained in S^*. Let P be the first point of AX_0, counting from A, which is a boundary point of S^*, and let B be any point of AX_0 beyond P such that PB is shorter than AP. We can now construct in S^* (compare the figure) two rectangles $ABCD$, $abcd$, which are concentric and coaxial. Their convex



closure is the octagon $A\ a\ b\ B\ D\ d\ c\ C$, of which P is an interior point. Therefore P cannot be a boundary point of S^*.

If $k \geqq 3$ we draw the same two-dimensional figure as before and rotate it around A by a small angle which moves the line AX_0 into a new position AX_0' nearer the point A_1. The two rectangles go over into new rectangles which are lying within S^* together with their boundaries, and, provided, the angle of rotation is small enough, the enclosing octagon will again contain the point P, although P is no longer located on the boundary of the rectangle $ABCD$. The two rectangles can be made the cross-sections of two k-dimensional rectangles whose remaining $k - 2$ dimensions are sufficiently small. In this way we can construct in S^* two k-dimensional rectangles which are concentric and coaxial and whose convex closure again contains the boundary point P.

REMARK. Our convex tube $\tilde{T}$ is a star region with respect to any of its points. In fact if it contains the points c_κ, z_κ, then, for any positive exponents σ_κ, the segment

$$\zeta_\kappa = c_\kappa + (z_\kappa - c_\kappa)t^{\sigma_\kappa}, \qquad 0 \leqq t \leqq 1,$$

is likewise contained in $\tilde{T}$. Hence, by a theorem of B. Almer,[7] if

$$f(z) = \sum a_{n_1 \cdots n_k}(z_1 - c_1)^{n_1} \cdots (z_k - c_k)^{n_k}$$

[6] Compare H. Tietze, *Über Konvexheit im kleinen und im grossen und über gewisse den Punkten einer Menge zugeordnete Dimensionszahlen*, Math. Zeitschrift, vol. 28 (1928), pp. 704–707.

[7] *Sur quelques problèmes de la théorie des fonctions analytiques de deux variables complexes.* Ark. Mat. Astron. Fys. vol. 17 (1922).

is the power series expansion of our function around the point c_κ, then for all points in $\tilde{T}$,

$$f(z) = \lim_{\epsilon \to 0} \sum \frac{a_{n_1 \cdots n_k}(z_1 - c_1)^{n_1} \cdots (z_k - c_k)^{n_k}}{\Gamma(1 + \epsilon\sigma_1 n_1 + \cdots + \epsilon\sigma_k n_k)},$$

In particular, since the point c_κ can be chosen as a point in S, we obtain a formula giving in one step the extension of our function $f(x)$ from the original domain S into its convex closure $\tilde{S}$.

PRINCETON UNIVERSITY.

Annals of Mathematics
Vol. 39, No. 1, January, 1938

EXTENSIONS OF LINEAR FUNCTIONALS, WITH APPLICATIONS TO LIMITS, INTEGRALS, MEASURES, AND DENSITIES[1]

By R. P. Agnew and A. P. Morse

(Received May 25, 1937)

1. **Introduction.** It is our intention to prove and give applications of a theorem (Theorem 2 below) related to Banach's theorem (Theorem 1) on extensions of linear functionals. The following definitions and conventions will materially shorten our work.

A function (or transformation) F with domain and range in vector spaces is called *linear* if

$$F(ax + by) = aF(x) + bF(y) \tag{1.1}$$

for all x and y in the domain of F and all real a and b. A function whose range is in the real number system is called a *functional*.

We shall use the symbol $[G, p, f, E]$ as an equivalent of the following statement:

E is a vector space; f is a linear functional defined over a linear manifold E_0 in E; G is a group of linear transformations which univalently map E into itself and E_0 into itself;[2] p is a functional with domain E such that

$$(1.2)\qquad \begin{aligned} p(x + y) &\leqq p(x) + p(y) && x, y \in E,\\ p(tx) &= tp(x) && t \geqq 0, x \in E,\\ p(g(x)) &= p(x) && g \in G, x \in E; \end{aligned}$$

and finally f and p are such that

$$(1.3)\qquad \begin{aligned} f(x) &\leqq p(x) && x \in E_0,\\ f(g(x)) &= f(x) && g \in G, x \in E_0. \end{aligned}$$

If $[G, p, f, E]$, then we use $\{G, p, f, E\}$ to denote the set of linear functionals F with domain E which satisfy

$$(1.4)\qquad \begin{aligned} F(x) &\leqq p(x) && x \in E,\\ F(x) &= f(x) && x \in E_0,\\ F(g(x)) &= F(x) && g \in G, x \in E. \end{aligned}$$

Thus if $F \in \{G, p, f, E\}$, then $F(x)$ is an extension with domain E of $f(x)$ which preserves for all $x \in E$ the properties ascribed by (1.3) to f.

---

[1] Presented to the American Mathematical Society September 7, 1937.

[2] If g_1, $g_2 \in G$, we use g_1g_2 to denote the product $h \in G$ defined by $h(x) = g_1(g_2(x))$ for $x \in E$.

We can now state Banach's theorem[3] (Theorem 1) in a manner which evidences the relation between it and our modification (Theorem 2) of it.

THEOREM 1. *If $[I, p, f, E]$, where I is the group of transformations of E into itself whose sole element is the identity transformation, then there exists a functional $F \epsilon \{I, p, f, E\}$.*

THEOREM 2.[4] *If $[G, p, f, E]$ and if there is a functional $F_1 \epsilon \{G^*, p, f, E\}$ where G^* is some derived group of G, then there exists a functional $F \epsilon \{G, p, f, E\}$.*

2. **Proof of Theorem 2.** If we denote the ν^{th} derived group of G by $G^{(\nu)}$, it becomes clear that Theorem 2 is implied by the following lemma.

LEMMA 2.01. *Let n be a positive integer. If $[G, p, f, E]$ and if there exists $F_1 \epsilon \{G^{(n)}, p, f, E\}$ then there exists $F \epsilon \{G^{(n-1)}, p, f, E\}$.*

To prove this lemma, let $H = G^{(n-1)}$, and define for each $x \epsilon E$

$$p_1(x) = \inf \sup_{h \epsilon H} \frac{1}{N} \sum_{j=1}^{N} F_1(hh_j(x)), \tag{2.1}$$

where the inf is taken over all positive integers N and all sequences $\{h_j\}$ in which $h_j \epsilon H$ for $j = 1, 2, \cdots$. Since the elements of G are linear transformations, i.e.

$$g(ax + by) = ag(x) + bg(y),$$

it is clear that

$$p_1(tx) = tp_1(x) \qquad t \geqq 0. \tag{2.2}$$

To show

$$p_1(x + y) \leqq p_1(x) + p_1(y) \qquad x, y \epsilon E, \tag{2.3}$$

let $\epsilon > 0$. Choose elements $h_1, h_2, \cdots, h_M$ and $h_1', h_2', \cdots, h_N'$ of H such that

$$\sup_{h \epsilon H} \frac{1}{M} \sum_{i=1}^{M} F_1(hh_i(x)) \leqq p_1(x) + \epsilon,$$

$$\sup_{h \epsilon H} \frac{1}{N} \sum_{j=1}^{N} F_1(hh_j'(y)) \leqq p_1(y) + \epsilon.$$

Then

$$\begin{aligned} p_1(x + y) &\leqq \sup_{h \epsilon H} \frac{1}{MN} \sum_{i,j=1}^{M,N} F_1(hh_ih_j'(x + y)) \\ &\leqq \sup_{h \epsilon H} \frac{1}{MN} \sum_{i,j=1}^{M,N} F_1(hh_ih_j'(x)) + \sup_{h \epsilon H} \frac{1}{MN} \sum_{i,j=1}^{M,N} F_1(hh_ih_j'(y)) \\ &\leqq \frac{1}{N} \sum_{j=1}^{N} \sup_{h \epsilon H} \frac{1}{M} \sum_{i=1}^{M} F_1(hh_ih_j'(x)) + \frac{1}{M} \sum_{i=1}^{M} \sup_{h \epsilon H} \frac{1}{N} \sum_{j=1}^{N} F_1(hh_ih_j'(y)). \end{aligned}$$

---

[3] Banach, Théorie des opérations linéaires, Warsaw (1932), pp. 27–29.

[4] The introduction of the general group G which appears here was suggested in part by J. von Neumann's paper: Zur allgemeinen Theorie des Masses, Funda. Math. Vol. 13 (1929) pp. 73–116.

Now if g_1, $g_2 \epsilon H$, then

$$F_1(g_1g_2(x)) = F_1[g_1g_2(g_2g_1)^{-1}g_2g_1(x)] = F_1(g_2g_1(x))$$

since the commutator $g_1g_2g_1^{-1}g_2^{-1} \equiv g_1g_2(g_2g_1)^{-1}$ is an element of the derived (or commutator) subgroup $H' \equiv G^{(n)}$ of $H \equiv G^{(n-1)}$ and under our hypotheses $F_1(g(x)) = F_1(x)$ for each $g \epsilon G^{(n)}$. Hence $F_1(hh_ih_j'(x)) = F_1(hh_j'h_i(x))$. Moreover if F^* is any functional with domain E and $g \epsilon H$, then, since H is a group,

$$\sup_{h \epsilon H} F^*(hg(x)) = \sup_{h \epsilon H} F^*(h(x)).$$

Hence we can continue the above estimate for $p_1(x + y)$ to obtain

$$p_1(x + y) \leqq \frac{1}{N}\sum_{j=1}^{N} \sup_{h \epsilon H} \frac{1}{M}\sum_{i=1}^{M} F_1(hh_i(x)) + \frac{1}{M}\sum_{i=1}^{M} \sup_{h \epsilon H} \frac{1}{N}\sum_{j=1}^{N} F_1(hh_j'(y))$$
$$\leqq p_1(x) + p_1(y) + 2\epsilon.$$

Arbitrariness of $\epsilon > 0$ establishes (2.3).

Now for x in the domain E_0 of f and $h \epsilon H$ we have $F_1(h(x)) = f(h(x)) = f(x)$ so that

$$p_1(x) \geqq \inf \frac{1}{N}\sum_{j=1}^{N} F_1(h_j(x)) = f(x), \tag{2.4}$$

the inf being taken as in (2.1). The properties (2.2), (2.3) and (2.4) enable us to apply Theorem 1 of Banach to show existence of a linear functional F with domain E for which

$$F(x) \leqq p_1(x) \leqq p(x) \qquad x \epsilon E,$$
$$F(x) = f(x). \qquad x \epsilon E_0.$$

If we show that F has also the property that

$$F(h_0(x)) = F(x) \qquad h_0 \epsilon H,\ x \epsilon E, \tag{2.5}$$

then we shall have $F \epsilon \{G^{(n-1)}, p, f, E\}$ and Lemma (2.01) will be proved. Since F is linear, it is sufficient to show that

$$F[h_0(x) - x] = 0 \qquad h_0 \epsilon H,\ x \epsilon E. \tag{2.6}$$

For each $N = 1, 2, 3, \cdots$ and $x \epsilon E$, we find

$$F[h_0(x) - x] \leqq p_1[h_0(x) - x] \leqq \sup_{h \epsilon H} \frac{1}{N}\sum_{j=1}^{N} F_1\{hh_0^j[h_0(x) - x]\}$$
$$= \sup_{h \epsilon H} \frac{1}{N}\sum_{j=1}^{N} [F_1(hh_0^{j+1}(x)) - F_1(hh_0^j(x))]$$
$$= \sup_{h \epsilon H} \frac{1}{N} [F_1(hh_0^{N+1}(x)) - F_1(hh_0(x))]$$

$$\leqq \sup_{h \in H} \frac{1}{N} [p(hh_0^{N+1}(x)) + p(hh_0(-x))]$$

$$= \sup_{h \in H} \frac{1}{N} [p(x) + p(-x)] = \frac{1}{N} [p(x) + p(-x)];$$

hence $F[h_0(x) - x] \leqq 0$. Since a similar argument shows that $-F[h_0(x) - x] = F[x - h_0(x)] \leqq 0$, we conclude validity of (2.6) and the proof of Lemma 2.01 is complete.

3. **Corollaries of Theorems 1 and 2.** As a corollary of Theorems 1 and 2, we have

THEOREM 3. *If* $[G, p, f, E]$ *with* G *a solvable group, then there exists a functional* $F \in \{G, p, f, E\}$.

We define $[G, p, E] = [G, p, f_0, E]$ and $\{G, p, E\} = \{G, p, f_0, E\}$ where the domain E_0 of the linear functional f_0 is the "zero" element θ of E. Observe that, for every linear functional f, $f(\theta) = 0$ and $\{G, p, f, E\} \subset \{G, p, E\}$. Identifying f with f_0 in Theorem 3 gives

THEOREM 4. *If* $[G, p, E]$ *with* G *a solvable group, then there exists a functional* $F \in \{G, p, E\}$.

4. **Generalized limits.** Let E be the vector space of real-valued functions $x \equiv x(s)$, defined over $-\infty < s < \infty$, for which $\overline{\lim}_{s\to\infty} | x(s) | < \infty$. Let $p(x) = \overline{\lim}_{s\to\infty} x(s)$. Let G be the group of transformations g from E to E such that for each $g \in G$,

$$\{g(x)\}(s) = x(\mu s + \lambda) \qquad x \in E, -\infty < s < \infty, \tag{4.1}$$

for some $\mu > 0$ and some real λ; for each $x \in E$ and each number s we employ the notation $\{g(x)\}(s)$ to denote the result of the function $g(x)$ operating on the number s. Clearly $[G, p, E]$. The commutator or derived group of G is the subgroup H of G for which $\mu = 1$; thus if $h \in H$, then

$$\{h(x)\}(s) = x(s + \lambda) \qquad x \in E, -\infty < s < \infty. \tag{4.2}$$

The group H being abelian, G is solvable and Theorem 4 assures existence of a functional $F \in \{G, p, E\}$.

Let E^* denote the set of functions x for which $x(s)$ is defined, real, and bounded for s sufficiently large. With each $x \in E^*$, associate a function $x^* \in E$ such that $x^*(s) = x(s)$ for all sufficiently large s. Upon defining $\mathrm{Lim}_{s\to\infty} x(s) = F(x^*)$, it may be seen that

$$\underline{\lim}_{s\to\infty} x(s) \leqq \mathrm{Lim}_{s\to\infty} x(s) \leqq \overline{\lim}_{s\to\infty} x(s), \tag{4.3}$$

$$\mathrm{Lim}_{s\to\infty} [ax(s) + by(s)] = a \,\mathrm{Lim}_{s\to\infty} x(s) + b \,\mathrm{Lim}_{s\to\infty} y(s), \tag{4.4}$$

$$\mathrm{Lim}_{s\to\infty} x(\mu s + \lambda) = \mathrm{Lim}_{s\to\infty} x(s), \tag{4.5}$$

for every $x, y \in E^*$ and set of real numbers a, b, μ, λ with $\mu > 0$.

It is easy to give bounded functions $x_1(s)$ and $x_2(s)$ for which[5]

$$\operatorname*{Lim}_{s\to\infty} x_1(s)x_2(s) \neq \operatorname*{Lim}_{s\to\infty} x_1(s) \operatorname*{Lim}_{s\to\infty} x_2(s);$$

however

$$\operatorname*{Lim}_{s\to\infty} x_1(s)x_2(s) = \lim_{s\to\infty} x_1(s) \operatorname*{Lim}_{s\to\infty} x_2(s) \tag{4.6}$$

whenever $\lim x_1(s)$ exists. To prove (4.6) write

$$x_1(s)x_2(s) = [x_1(s) - \lim_{s\to\infty} x_1(s)]x_2(s) + [\lim_{s\to\infty} x_1(s)]x_2(s)$$

and apply (4.4) and (4.3).

5. **Generalized Lebesgue integrals.** Let E be the vector space of real-valued functions $x \equiv x(s)$ defined over $-\infty < s < \infty$ for which $\int^* | x(s) | \, ds < \infty$ where $\int^* x(s)\, ds$ denotes the upper Lebesgue integral over $-\infty < s < \infty$ of $x(s)$. Observe that E contains as a vector sub-space the class of all functions $x \equiv x(s)$ each of which is real and bounded over $-\infty < s < \infty$ and vanishes outside some finite interval of values of s. Define $p(x) = \int^* x(s)\, ds$ for $x \in E$. Let G be the group of transformations g from E to E for which

$$\{g(x)\}(s) = | \mu | \, x(\mu s + \lambda) \qquad x \in E, \; -\infty < s < \infty, \tag{5.1}$$

for some real μ and λ with $\mu \neq 0$. Again, $[G, p, E]$; and the commutator or derived group of G is the abelian subgroup H consisting of the elements of G for which $\mu = 1$. Hence Theorem 4 furnishes $F \in \{G, p, E\}$.

Upon defining

$$^{(1)}\!\!\int_{-\infty}^{\infty} x(s)\, ds = F(x), \tag{5.2}$$

we have an integral, defined for each $x \in E$, having the following properties:

$$\int_* x(s)\, ds \leqq {}^{(1)}\!\!\int_{-\infty}^{\infty} x(s)\, ds \leqq \int^* x(s)\, ds, \tag{5.3}$$

$$^{(1)}\!\!\int_{-\infty}^{\infty} [ax(s) + by(s)]\, ds = a \, {}^{(1)}\!\!\int_{-\infty}^{\infty} x(s)\, ds + b \, {}^{(1)}\!\!\int_{-\infty}^{\infty} y(s)\, ds, \tag{5.4}$$

$$^{(1)}\!\!\int_{-\infty}^{\infty} | \mu | \, x(\mu s + \lambda)\, ds = {}^{(1)}\!\!\int_{-\infty}^{\infty} x(s)\, ds, \tag{5.5}$$

---

[5] For example, let $x_1(s) = x_2(s) = \cos s$. Then

$$\operatorname{Lim} x_1(s) = \tfrac{1}{2} \operatorname{Lim} [x_1(s) + x_1(s + \pi)] = \tfrac{1}{2} \operatorname{Lim} 0 = 0 \text{ while}$$

$$\operatorname{Lim} [x_1(s)]^2 = \operatorname{Lim} \tfrac{1}{2}[1 + x_1(2s)] = \tfrac{1}{2}.$$

where $\int_*$ and $\int^*$ represent lower and upper Lebesgue integrals; $x, y \in E$; and a, b, μ, λ are real constants with $\mu \neq 0$.

Letting $\varphi_A(s)$ represent the characteristic function of a set A, we complete our definition of $\overset{(1)}{\int}$ by defining

$$\overset{(1)}{\int_A} x(s)\,ds = \overset{(1)}{\int_{-\infty}^{\infty}} x(s)\,\varphi_A(s)\,ds \tag{5.6}$$

whenever the right member exists. If A is an interval (a, b), we write the integral in the conventional form

$$\overset{(1)}{\int_a^b} x(s)\,ds. \tag{5.7}$$

It is easy to modify formulas (5.3), (5.4), and (5.5) to obtain properties of the more general integral (5.6). We observe that if $x(s)$ is integrable over A, then it is integrable over any subset of A; and that if $A = A_1 + A_2$ and the intersection A_1A_2 is null (i.e. A_1A_2 has Lebesgue measure $|A_1A_2| = 0$) then

$$\overset{(1)}{\int_{A_1}} x(s)\,ds + \overset{(1)}{\int_{A_2}} x(s)\,ds = \overset{(1)}{\int_A} x(s)\,ds. \tag{5.8}$$

We point out finally that if A_1 is a measurable set containing A, and $x_1(s)$ is a function for which $|x(s)| \leqq x_1(s)$ for $s \in A$ and the Lebesgue integral

$$\overset{(L)}{\int_{A_1}} x_1(s)\,ds \qquad \text{exists, then} \qquad \overset{(1)}{\int_A} x(s)\,ds$$

exists; in particular, every bounded function is integrable over every bounded set.

6. **Generalized Lebesgue measure.** Let D be the class of point sets in the space $-\infty < s < \infty$ having finite upper Lebesgue measure. For each $A \in D$, let $\varphi_A(s)$ be the characteristic function of A, and let

$$m_1(A) = \overset{(1)}{\int_{-\infty}^{\infty}} \varphi_A(s)\,ds \equiv \overset{(1)}{\int_A} 1\,ds. \tag{6.1}$$

We now have a measure-function (or set-function) $m_1(A)$ defined for all $A \in D$ and having the following properties: for each $A \in D$,

$$m_*(A) \leqq m_1(A) \leqq m^*(A) \tag{6.2}$$

where m_* and m^* denote lower and upper Lebesgue measures; if $A, B \in D$ and the intersection AB is empty, then

$$m_1(A + B) = m_1(A) + m_1(B); \tag{6.3}$$

and if $A \in D$, μ and λ are real constants, and $A_{\mu,\lambda}$ is the set of points s' representable in the form $\mu s + \lambda$ with $s \in A$, then

$$m_1(A_{\mu,\lambda}) = |\mu|\, m_1(A). \tag{6.4}$$

7. **Generalized Riemann integrals.** Let ${}^{(R)}\int_a^b \xi(s)\,ds$ denote the upper Darboux (sometimes called upper Riemann) integral of a real bounded function $\xi(s)$ over a finite interval $a \leqq s \leqq b$. A real function $x(s)$ being given, let $x_{\alpha,\beta}(s)$ denote the "chopped-off" function defined by $x_{\alpha,\beta}(s) = \alpha,\ x(s)$, or β according as $x(s) \leqq \alpha,\ \alpha < x(s) < \beta$, or $\beta \leqq x(s)$. Let

$$(7.1)\qquad {}^{(R)}\int^{*} x(s)\,ds = \lim_{-a,b,-\alpha,\beta\to\infty} {}^{(R)}\int_a^b x_{\alpha,\beta}(s)\,ds$$

when the limit exists. Similarly, let

$$(7.2)\qquad {}_{(R)}\int_{*} x(s)\,ds = \lim_{-a,b,-\alpha,\beta\to\infty} {}_{(R)}\int_a^b x_{\alpha,\beta}(s)\,ds$$

when the limit exists and ${}_{(R)}\int_a^b$ is a lower Darboux integral.

Let E be the vector space of real functions $x \equiv x(s)$, defined over $-\infty < s < \infty$, for which

$$(7.3)\qquad {}^{(R)}\int_{*} |\, x(s) \,|\, ds < \infty,$$

and let

$$(7.4)\qquad p(x) = {}^{(R)}\int^{*} x(s)\,ds \qquad\qquad x \in E.$$

With the solvable group G defined precisely as in §5, we verify that $[G, p, E]$. Hence we can use Theorem 4 to obtain existence (but not uniqueness) of $F \in \{G, p, E\}$ which yields an integral

$$(7.5)\qquad {}^{(2)}\int_{-\infty}^{\infty} x(s)\,ds = F(x) \qquad\qquad x \in E$$

which might or might not be consistent[6] with the integral of §5. Properties of ${}^{(2)}\int$ can be developed as properties of ${}^{(1)}\int$ were developed in §5.

We now proceed to define an integral, ${}^{(3)}\int$, which belongs to a special class of generalized Riemann integrals which are necessarily inconsistent with ${}^{(1)}\int$ and the Lebesgue integral. The range of application of ${}^{(3)}\int$ will be larger than the range for ${}^{(2)}\int$, and will include functions not in the range for ${}^{(1)}\int$. We call a

[6] We say that two integrals ${}^{(1)}\int$ and ${}^{(2)}\int$ are consistent if existence of both ${}^{(1)}\int x(s)\,ds$ and ${}^{(2)}\int x(s)\,ds$ implies their equality.

set C of real numbers *essentially non-dense* if $C = A + B$ where A is non-dense in one-space and B has Lebesgue measure $|B| = 0$. Let K be the class of functions $k \equiv k(s)$ each of which vanishes for all s except those of an essentially non-dense set. Let $\tilde{E}$ be the linear space of functions $x \equiv x(s)$ for which

$$\inf_{k \in K} {}^{(R)}\!\int^{*} |x(s) - k(s)|\, ds < \infty.$$

For each $x \in \tilde{E}$, let

$$\tilde{p}(x) = \inf_{k \in K} {}^{(R)}\!\int^{*} [x(s) - k(s)]\, ds.$$

Again using the group G of §5, we verify that $[G, p, \tilde{E}]$ and obtain $F \in \{G, p, \tilde{E}\}$. Upon defining

$${}^{(3)}\!\int_{-\infty}^{\infty} x(s)\, ds = \tilde{F}(x),$$

we have an integral with the following properties: for each $x, y \in \tilde{E}$ and set of real constants a, b, μ, λ, with $\mu \neq 0$,

$$\text{(7.6)} \quad \sup_{k \in K} {}_{(R)}\!\int_{*} [x(s) + k(s)]\, ds \leqq {}^{(3)}\!\int_{-\infty}^{\infty} x(s)\, ds \leqq \inf_{k \in K} {}^{(R)}\!\int^{*} [x(s) - k(s)]\, ds;$$

$$\text{(7.61)} \quad {}^{(3)}\!\int_{-\infty}^{\infty} [ax(s) + by(s)]\, ds = a\, {}^{(3)}\!\int_{-\infty}^{\infty} x(s)\, ds + b\, {}^{(3)}\!\int_{-\infty}^{\infty} y(s)\, ds;$$

$$\text{(7.62)} \quad {}^{(3)}\!\int_{-\infty}^{\infty} |\mu|\, x(\mu s + \lambda)\, ds = {}^{(3)}\!\int_{-\infty}^{\infty} x(s)\, ds;$$

$$\text{(7.63)} \quad {}^{(3)}\!\int x_0(s)\, ds = 0, \qquad x_0 \in K.$$

The property (7.6) is merely an expression of the inequality $-\tilde{p}(-x) \leqq \tilde{F}(x) \leqq \tilde{p}(x)$. Proof of (7.63) consists in noting that if $x_0 \in K$, then $-\tilde{p}(-x_0) = \tilde{p}(x_0) = 0$. It follows from (7.6) that if $x \in \tilde{E}$, then

$$\text{(7.64)} \quad {}_{(R)}\!\int_{*} x(s)\, ds \leqq {}^{(3)}\!\int_{-\infty}^{\infty} x(s)\, ds \leqq {}^{(R)}\!\int^{*} x(s)\, ds$$

whenever the extreme integrals exist. The space E of this section is a subspace of $\tilde{E}$, so (7.64) holds in particular for each $x \in E$; thus ${}^{(3)}\!\int$ is a generalized Riemann integral.

We complete the definition of ${}^{(3)}\!\int$ by defining

$$\text{(7.65)} \quad {}^{(3)}\!\int_{A} x(s)\, ds = {}^{(3)}\!\int_{-\infty}^{\infty} x(s)\, \varphi_A(s)\, ds$$

and can obtain properties of this more general integral by paralleling work of §5.

We notice that if $x(s)$ vanishes for all s except those of a null set (set of Lebesgue measure 0) then $\overset{(3)}{\int_{-\infty}^{\infty}} x(s)\, ds = 0$. However there exist many Lebesgue integrable functions $x(s)$ (in fact, many which are bounded and vanish outside a finite interval) for which

$$\overset{(3)}{\int_{-\infty}^{\infty}} x(s)\, ds \neq \overset{(1)}{\int_{-\infty}^{\infty}} x(s)\, ds = \overset{(L)}{\int_{-\infty}^{\infty}} x(s)\, ds.$$

For example, let $x_0(s)$ be the characteristic function of a bounded Lebesgue measurable essentially non-dense set A_0 of measure $|A_0| = 1$. Then $x_0 \in K$ and

$$0 = \overset{(3)}{\int_{-\infty}^{\infty}} x_0(s)\, ds \neq \overset{(1)}{\int_{-\infty}^{\infty}} x_0(s)\, ds = |A_0| = 1. \tag{7.7}$$

This set A_0 may be taken to be the complement in the interval $0 \leqq s \leqq 2$ of the union $B_0 = I_1 + I_2 + \cdots$ of a countable set of mutually exclusive open subintervals I_n of $(0, 2)$. If we put $B_n = I_1 + I_2 + \cdots + I_n$ and let $y_n(t)$ be the characteristic function of B_n, then

$$\lim_{n\to\infty} \overset{(3)}{\int_0^2} y_n(s)\, ds = \lim_{n\to\infty} \sum_{k=1}^{n} |I_k| = 1; \tag{7.8}$$

however $y_0(s) = \lim_{n\to\infty} y_n(s)$ is the characteristic function of B_0 and

$$\overset{(3)}{\int_0^2} y_0(s)\, ds = \overset{(3)}{\int_0^2} [y_0(s) + x_0(s)]\, ds - \overset{(3)}{\int_0^2} x_0(s)\, ds = 2 - 0 = 2. \tag{7.9}$$

Formulas (7.7), (7.8), and (7.9) show that the special extension $\overset{(3)}{\int}$ of the Riemann integral is not only inconsistent with the Lebesgue integral but also fails to possess certain fundamental properties of the Lebesgue integral.

8. **Generalized Jordan Content.** Let D be the class of point sets A in one-space having finite upper Jordan content $M^*(A)$. For each $A \in D$ let $\varphi_A(s)$ be the characteristic function of A and let

$$m_2(A) = \overset{(2)}{\int_{-\infty}^{\infty}} \varphi_A(s)\, ds \equiv \overset{(2)}{\int_A} 1\, ds. \tag{8.1}$$

Leaving discussion of the measure function (or set function) $m_2(A)$ to the reader, we proceed to use $\overset{(3)}{\int}$ to define a measure function $m_3(A)$ which belongs to a special class of generalized Jordan contents which are necessarily inconsistent with m_1 and Lebesgue measure. The range of application of m_3 will be larger than the range for m_2.

Let $\mathfrak{N}$ be the class of essentially non-dense sets N, and let $\tilde{D}$ be the class of sets A for which

$$\inf_{N \in \mathfrak{N}} M^*(A - N) < \infty. \tag{8.2}$$

For each $A \in \tilde{D}$, the integrals in (8.3) exist and we define $m_3(A)$ by

$$m_3(A) = {}^{(3)}\!\!\int_{-\infty}^{\infty} \varphi_A(s)\, ds \equiv {}^{(3)}\!\!\int_A 1\, ds, \qquad A \in \tilde{D}. \tag{8.3}$$

Using $M_*(A)$ amd $M^*(A)$ to denote lower and upper Jordan contents of A, we see that $m_3(A)$ has the following properties: if $A \in \tilde{D}$, then

$$\sup_{N \in \mathfrak{N}} M^*(A + N) \leqq m_3(A) \leqq \inf_{N \in \mathfrak{N}} M_*(A - N); \tag{8.4}$$

if A, $B \in \tilde{D}$ and AB is empty, then

$$m_3(A + B) = m_3(A) + m_3(B); \tag{8.41}$$

if $A \in \tilde{D}$, μ and λ are real constants, and $A_{\mu,\lambda}$ is the set of points s' representable in the form $\mu s + \lambda$ with $s \in A$, then

$$m_3(A_{\mu,\lambda}) = |\mu|\, m_3(A); \tag{8.42}$$

and if N is essentially non-dense, then

$$m_3(N) = 0. \tag{8.43}$$

In particular if $A \in D$, then $A \in \tilde{D}$; hence we see from (8.4) that

$$M_*(A) \leqq m_3(A) \leqq M^*(A) \tag{8.44}$$

whenever the upper Jordan content of A is finite and $m_3(A)$ is a generalized Jordan content.

With the mutually exclusive open intervals $I_1, I_2, \cdots$ defined in §7, we have

$$2 = m_3(I_1 + I_2 + \cdots) \neq m_3(I_1) + m_3(I_2) + \cdots = 1; \tag{8.5}$$

hence m_3 is not a completely additive set function and

$$m_3(I_1 + I_2 + \cdots) \neq m_1(I_1 + I_2 + \cdots) = |I_1 + I_2 + \cdots|. \tag{8.6}$$

This measure function settles the interesting question whether a measure function $m(A)$ can satisfy (8.44), (8.41), (8.42), and the condition that $m(A) = 0$ for each null set A, and nevertheless fail to be completely additive.

9. **Generalized density of point sets.** Let A be a set and s a point in the interval $-\infty < s < \infty$. Let $A_s(t)$ be the subset of A in the interval $I(s, t)$ with center at s and length t^{-1}. Then $m_1(A_s(t))/t^{-1}$ is a generalized average density of A over the interval $I(s, t)$ and

$$D(A, s) = \operatorname*{Lim}_{t \to \infty} m_1(A_s(t))/t^{-1}$$

We notice that if $x(s)$ vanishes for all s except those of a null set (set of Lebesgue measure 0) then $\overset{(3)}{\int}_{-\infty}^{\infty} x(s)\, ds = 0$. However there exist many Lebesgue integrable functions $x(s)$ (in fact, many which are bounded and vanish outside a finite interval) for which

$$\overset{(3)}{\int}_{-\infty}^{\infty} x(s)\, ds \neq \overset{(1)}{\int}_{-\infty}^{\infty} x(s)\, ds = \overset{(L)}{\int}_{-\infty}^{\infty} x(s)\, ds.$$

For example, let $x_0(s)$ be the characteristic function of a bounded Lebesgue measurable essentially non-dense set A_0 of measure $|A_0| = 1$. Then $x_0 \epsilon K$ and

$$0 = \overset{(3)}{\int}_{-\infty}^{\infty} x_0(s)\, ds \neq \overset{(1)}{\int}_{-\infty}^{\infty} x_0(s)\, ds = |A_0| = 1. \tag{7.7}$$

This set A_0 may be taken to be the complement in the interval $0 \leqq s \leqq 2$ of the union $B_0 = I_1 + I_2 + \cdots$ of a countable set of mutually exclusive open subintervals I_n of $(0, 2)$. If we put $B_n = I_1 + I_2 + \cdots + I_n$ and let $y_n(t)$ be the characteristic function of B_n, then

$$\lim_{n\to\infty} \overset{(3)}{\int}_{0}^{2} y_n(s)\, ds = \lim_{n\to\infty} \sum_{k=1}^{n} |I_k| = 1; \tag{7.8}$$

however $y_0(s) = \lim_{n\to\infty} y_n(s)$ is the characteristic function of B_0 and

$$\overset{(3)}{\int}_{0}^{2} y_0(s)\, ds = \overset{(3)}{\int}_{0}^{2} [y_0(s) + x_0(s)]\, ds - \overset{(3)}{\int}_{0}^{2} x_0(s)\, ds = 2 - 0 = 2. \tag{7.9}$$

Formulas (7.7), (7.8), and (7.9) show that the special extension $\overset{(3)}{\int}$ of the Riemann integral is not only inconsistent with the Lebesgue integral but also fails to possess certain fundamental properties of the Lebesgue integral.

8. **Generalized Jordan Content.** Let D be the class of point sets A in one-space having finite upper Jordan content $M^*(A)$. For each $A \epsilon D$ let $\varphi_A(s)$ be the characteristic function of A and let

$$m_2(A) = \overset{(2)}{\int}_{-\infty}^{\infty} \varphi_A(s)\, ds \equiv \overset{(2)}{\int}_{A} 1\, ds. \tag{8.1}$$

Leaving discussion of the measure function (or set function) $m_2(A)$ to the reader, we proceed to use $\overset{(3)}{\int}$ to define a measure function $m_3(A)$ which belongs to a special class of generalized Jordan contents which are necessarily inconsistent with m_1 and Lebesgue measure. The range of application of m_3 will be larger than the range for m_2.

Let $\mathfrak{N}$ be the class of essentially non-dense sets N, and let $\tilde{D}$ be the class of sets A for which

$$\inf_{N \in \mathfrak{N}} M^*(A - N) < \infty. \tag{8.2}$$

For each $A \in \tilde{D}$, the integrals in (8.3) exist and we define $m_3(A)$ by

$$m_3(A) = {}^{(3)}\!\!\int_{-\infty}^{\infty} \varphi_A(s)\, ds \equiv {}^{(3)}\!\!\int_A 1\, ds, \qquad A \in \tilde{D}. \tag{8.3}$$

Using $M_*(A)$ amd $M^*(A)$ to denote lower and upper Jordan contents of A, we see that $m_3(A)$ has the following properties: if $A \in \tilde{D}$, then

$$\sup_{N \in \mathfrak{N}} M^*(A + N) \leqq m_3(A) \leqq \inf_{N \in \mathfrak{N}} M_*(A - N); \tag{8.4}$$

if A, $B \in \tilde{D}$ and AB is empty, then

$$m_3(A + B) = m_3(A) + m_3(B); \tag{8.41}$$

if $A \in \tilde{D}$, μ and λ are real constants, and $A_{\mu,\lambda}$ is the set of points s' representable in the form $\mu s + \lambda$ with $s \in A$, then

$$m_3(A_{\mu,\lambda}) = |\mu|\, m_3(A); \tag{8.42}$$

and if N is essentially non-dense, then

$$m_3(N) = 0. \tag{8.43}$$

In particular if $A \in D$, then $A \in \tilde{D}$; hence we see from (8.4) that

$$M_*(A) \leqq m_3(A) \leqq M^*(A) \tag{8.44}$$

whenever the upper Jordan content of A is finite and $m_3(A)$ is a generalized Jordan content.

With the mutually exclusive open intervals $I_1, I_2, \cdots$ defined in §7, we have

$$2 = m_3(I_1 + I_2 + \cdots) \neq m_3(I_1) + m_3(I_2) + \cdots = 1; \tag{8.5}$$

hence m_3 is not a completely additive set function and

$$m_3(I_1 + I_2 + \cdots) \neq m_1(I_1 + I_2 + \cdots) = |I_1 + I_2 + \cdots|. \tag{8.6}$$

This measure function settles the interesting question whether a measure function $m(A)$ can satisfy (8.44), (8.41), (8.42), and the condition that $m(A) = 0$ for each null set A, and nevertheless fail to be completely additive.

9. **Generalized density of point sets.** Let A be a set and s a point in the interval $-\infty < s < \infty$. Let $A_s(t)$ be the subset of A in the interval $I(s, t)$ with center at s and length t^{-1}. Then $m_1(A_s(t))/t^{-1}$ is a generalized average density of A over the interval $I(s, t)$ and

$$D(A, s) = \operatorname*{Lim}_{t \to \infty} m_1(A_s(t))/t^{-1}$$

is a generalized density of A at the point s. Formulation of properties of this density is left to the reader. We can form other density functions by using m_2 or m_3 instead of m_1. In any of these cases, every set has a density at every point.

10. **Densities of sets of positive integers.** Let A be a set of positive integers, and let $N(A, s)$ be the number of integers in the set A which are less than or equal to s. We define the *density* $D(A)$ of the set A by the formula

$$D(A) = \operatorname*{Lim}_{s \to \infty} N(A, s)/s.$$

The inequality $0 \leqq N(A, s)/s \leqq 1$, which holds for each $s > 0$, insures that each set A of positive integers has a density. For example, the set E of even integers has density $D(E) = \frac{1}{2}$, the set P of primes has density $D(P) = 0$, and the set C of composite numbers has density $D(C) = 1$. This density has the property that $0 \leqq D(A) \leqq 1$ for each set A of positive integers; that

$$D(A + B) = D(A) + D(B)$$

when A and B are sets without common elements; and that if the elements of A are represented as the elements of a sequence $\{a_n\}$, then

$$D(\{\mu a_n + \lambda\}) = D(\{a_n\})/\mu$$

for each pair μ and λ of integers with $\mu > 0$, $\lambda \geqq 0$.

CORNELL UNIVERSITY, ITHACA, N. Y.
BROWN UNIVERSITY, PROVIDENCE, R. I.

Annals of Mathematics
Vol. 39, No. 1, January, 1938

GREEN'S FORMULAS FOR ANALYTIC FUNCTIONS[1]

By Hillel Poritsky

(Received May 25, 1937)

1. **Introduction.** In the complex plane one can express an analytic function over a region in terms of its boundary values by means of Cauchy's integral. In a similar way Green's formula furnishes a way of expressing a harmonic function in terms of the function and its normal derivative on the boundary. The latter formula for three dimensions runs

$$-u_P = \frac{1}{4\pi}\int\left(u\,\frac{\partial(1/r)}{\partial n} - \frac{\partial u}{\partial n}\frac{1}{r}\right)dS.$$

A similar result holds for n dimensions if 4π is replaced by K_n, the $(n - 1)$-dimensional area of a unit sphere in n dimensions, and $1/r$ by $r^{2-n}/(n - 2)$ if $n > 2$, or by $-\log r$ if $n = 2$.

In this paper is established a generalized Green's formula which expresses any analytic function u in terms of boundary values of the function and certain of its derivatives. For $n = 3$ this formula is as follows:

$$-4\pi u_P = \int\Bigg[\left(u\,\frac{\partial(1/r)}{\partial n} - \frac{\partial u}{\partial n}\frac{1}{r}\right) + \left(\nabla^2 u\,\frac{\partial(r/2!)}{\partial n} - \frac{\partial(\nabla^2 u)}{\partial n}\frac{r}{2!}\right) + \left(\nabla^4 u\,\frac{\partial(r^3/4!)}{\partial n} - \frac{\partial(\nabla^4 u)}{\partial n}\frac{r^3}{4!}\right) + \cdots\Bigg]ds.$$

For harmonic functions u this reduces to the Green's formula.

The method of proof is analogous to the method of establishing Taylor's expansion by successive integration by parts, and for one dimension ($n = 1$) actually yields a result equivalent to Taylor's expansion. A finite series with a remainder is derived (§2), and the latter is shown to approach zero under proper conditions. For arbitrary analytic functions the result is established for a sufficiently small region (§4). For certain functions, however, the result is valid for arbitrary regions (§5); this is particularly the case for functions satisfying the differential equation

$$\text{II (13)} \qquad \sum_{i=0}^{p} c_i \nabla^{2i} u = 0,$$

[1] Presented to the Society December 1928 under the title "Cauchy-Green Expansions." This paper constitutes essentially the second chapter of the Author's Ph.D. dissertation, Cornell, 1927, "Topics in Potential Theory." The material of the first chapter has appeared in somewhat expanded form in the papers cited as I and II below.

in which case the sum of the infinite series is given in finite form. Various interesting applications of the above results are given in §6. The expansions in question are used to show the analyticity of solutions of II (13).

The notation "II (13)" pertains to eq. (13) in a paper "Generalizations of the Gauss Law of the Mean"[2] to which we shall refer briefly as II. The present paper is essentially independent of II; it has, however, several points of contact with it, and to save space references are made to II freely. In particular the *symmetric* solutions of II (13), that is the solutions depending on r, the distance from a fixed point, are taken from II (II, §2) without any further discussion, while the solution

$$\text{II (36)} \qquad \Psi_n(r, \lambda) = \sum_{k=0}^{\infty} \lambda^k V_{k+1}(r)$$

of the special case of II (13):

$$\text{II (1)} \qquad \nabla^2 u - \lambda u = 0, \qquad \lambda = \text{const.},$$

which is singular at $r = 0$ is utilized extensively in the present paper.

In the same manner as Green's formula leads to the expansions here considered, the expression of a harmonic function in terms of its boundary values only, by means of Green's function, suggests a similar expression of an arbitrary analytic function u in terms of boundary values of u, $\nabla^2 u$, $\nabla^4 u$, $\cdots$ (without bringing in the normal derivatives of u, $\nabla^2 u$, $\cdots$). This can be done by introducing proper Green's functions. The resulting expansion has been treated briefly in the last section of the paper "On Certain Polynomial and Other Approximations to Analytic Functions."[3] The convergence of the resulting series is, however, of an entirely different nature from the one considered below.

2. The Generalized Green's Formulas with and without the Remainder. Consider Green's theorem applied successively to pairs of functions

$$u_1, v_1; u_2, v_2; \cdots; u_m, v_m:$$

$$(1) \qquad \begin{cases} \int (u_1 \nabla^2 v_1 - v_1 \nabla^2 u_1)\, dt = \int (u_1 \partial v_1/\partial n - v_1 \partial u_1/\partial n)\, dS, \\ \int (u_2 \nabla^2 v_2 - v_2 \nabla^2 u_2)\, dt = \int (u_2 \partial v_2/\partial n - v_2 \partial u_2/\partial n)\, dS, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \int (u_m \nabla^2 v_m - v_m \nabla^2 u_m)\, dt = \int (u_m \partial v_m/\partial n - v_m \partial u_m/\partial n)\, dS. \end{cases}$$

Here S is a closed $(n - 1)$-dimensional surface in n dimensions bounding the volume t; u_i, v_i are to possess continuous second derivatives. Now choose u_1 an arbitrary analytic function u and let

$$(2) \qquad u_2 = \nabla^2 u_1, \qquad u_3 = \nabla^2 u_2, \cdots, \qquad u_m = \nabla^2 u_{m-1};$$

[2] To appear soon in the Transactions of the American Mathematical Society. The paper II is a sequel to I: "On Operations Permutable with the Laplacian," American Journal of Mathematics, vol. 54, (1932), p. 667.

[3] Transactions of the American Mathematical Society, vol. 34, pp. 274–331.

also choose v_i so as to satisfy the relations

$$\nabla^2 v_1 = 0, \qquad \nabla^2 v_2 = v_1, \cdots, \qquad \nabla^2 v_m = v_{m-1}. \tag{3}$$

Adding equations (1) we get

$$-\int v_m \nabla^{2n} u \, dt = \sum_{i=1}^{m} \int (u_i \, \partial v_i/\partial n - v_i \, \partial u_i/\partial n) \, dS. \tag{4}$$

Let P be an arbitrary point inside S, r the distance from it. We choose

$$v_i = V_i(r)$$

where $V_i(r)$ are the functions referred to in the preceding section. These functions depend on r only; in particular,

$$V_1(r) = \begin{cases} \log r & \text{if } n = 2, \\ r^{2-n}/(2 - n) & \text{if } n \neq 2, \end{cases} \tag{5_1}$$

while in general

$$V_i(r) = r^{(2i-n)}(a_{n,i} + b_{n,i} \log r), \tag{5_i}$$

where $a_{n,i}$, $b_{n,i}$ are proper constants. For odd n, $b_{n,i}$ vanish. As explained in §1, the functions V_i were introduced in II by means of the expansion $\sum_{k=0}^{\infty} \lambda^k V_{k+1}(r)$, which converges for all λ and $r \neq 0$ to a function $\Psi_n(r, \lambda)$ satisfying II (1) except at $r = 0$. That the functions $V_i(r)$ will then satisfy (3) except at $r = 0$ is shown in II, §2.

In view of the singular nature of $P(r = 0)$ we exclude the neighborhood of the point P in applying (4), by restricting the volume t to the volume t' inside S and outside a small sphere Σ whose radius is ϵ and whose center is at P. The surface integrations have to be extended over both S and Σ; along the latter $\partial/\partial n$ may be replaced by $-\partial/\partial r$ or by the negative of the derivative along the radius. We proceed to show that the surface integrals along Σ approach zero with ϵ except for $\int u \, \partial v_1/\partial n \, d\Sigma$, and that the latter approaches $K_n u_P$.

From (5) follows that as ϵ approaches zero the behavior of $V_i(r)$ over Σ may be described by

$$V_i(\epsilon) = O(\epsilon^{2i-n} \log r).$$

Since u_i, $\partial u_i/\partial n$ are bounded (for any one i),

$$\int V_i \partial u_i/\partial n \, d\Sigma = O(\epsilon^{2i-1} \log \epsilon);$$

there integrals, consequently, all approach zero with ϵ. Again,

$$\partial V_i/\partial r = O(\epsilon^{2i-n-1} \log \epsilon);$$

hence

$$\int u_i \, \partial V_i/\partial n \, d\Sigma = O(\epsilon^{2i} \log \epsilon) = o(1), \qquad \text{if } i > 1.$$

For $i = 1$ $\partial V_1/\partial r = r^{1-n}$, and the last integral is equal to

$$-\epsilon^{1-n}\int u_1\,d\Sigma = -\int u_1\,d\omega,$$

where $d\omega$ is the area of the projection of $d\Sigma$ from the center P on the unit concentric sphere, but where u_1 is evaluated over the original Σ. Therefore

$$\lim_{\epsilon\to 0}\int u_1 \partial V_1/\partial n\, d\Sigma = -u_P K_n.$$

Carrying these results back to (4), there follows

$$\text{(6)}\quad \begin{aligned} K_n u_P &= \sum_{i=1}^{m}\int (u_i \partial V_i/\partial n - V_i\,\partial u_i/\partial n)\,ds + \int V_m \nabla^2 u_m\,dt \\ &= \sum_{i=1}^{m}\int [\nabla^{2i-2}u\,\partial V_i/\partial n - V_i\,\partial(\nabla^{2i-2}u)/\partial n]\,ds + \int V_m \nabla^{2m} u\,dt. \end{aligned}$$

Denote the volume integral above by R_m. If R_m approaches zero as m becomes infinite, then

$$\text{(7)}\quad K_n u_P = \sum_{i=1}^{\infty}\int [\nabla^{2i-2}u\,\partial V_i/\partial n - V_i\,\partial(\nabla^{2i-2}u)/\partial n]\,dS.$$

If the point P is taken outside of the surface S we may apply (4) at once (without introducing the sphere Σ) and find that the last member of (6) is equal to zero. Finally, if P is on S, its neighborhood has to be excluded again; we now obtain as a coefficient of u_P in (6) (under proper restrictions on S) $K_n/2$.

3. **Interpretation of (6) and (7) from Point of View of Potential Theory.** The formulas derived in last section enable one to regard the function u as the potential of certain central forces emanating from elements spread over S and t. Thus, if, to take a concrete example, we put $n = 3$, then $V_1(r) = -1/r$, $V_2(r) = -r/2!$, $V_3(r) = -r^3/4!$, $\cdots$. Now $V_1(r)$ becomes the (Newtonian) potential of a central force that varies according to the inverse square law;

$$\int (u\,\partial V_1/\partial n - V_1\,\partial u/\partial n)\,dS = -\int (u\partial(1/r)/\partial n - (1/r)\partial u/\partial n]\,dS$$

consequently represents the potential of the familiar Green *double layer* consisting of a layer of poles over S of density $-\partial u/\partial n$ and of a layer of normal doublets or dipoles of moment u.

In a similar way we may interpret the term $\int [\nabla^2 u\,\partial V_2/\partial n - V_2\,\partial(\nabla^2 u)/\partial n]\,dS$ of (6) or (7) as the potential of a layer of poles and doublets over S of densities $-\partial(\nabla^2 u)/\partial n$ and $\nabla^2 u$, respectively, where the elements exert central forces of magnitude $dV_2(r)/dr$. If $n = 3$, this magnitude is constant. Similar considerations apply to the other surface integrals of (6) or (7). As to the volume integral of (6), it represents the potential of elements spread over the volume t of volume density $\nabla^{2m}u$ and exerting central forces possessing the potential $V_m(r)$. Thus any analytic function can be considered as the potential of the

fields just described, barring for the moment questions of the convergence of the remainder in (7) to zero. Equation (6) with its above interpretation obviously applies to functions u that are not analytic, but merely possess a proper number of continuous derivatives.

It is readily seen that $V_i(r)$ is *i-harmonic*;[4] more precisely, if $r = PP'$, $V_i(r)$ is *i-harmonic* in the coordinates of P except at P'. It follows from this that the several surface integrals of (6) represent functions that are respectively *harmonic, bi-harmonic,* $\cdots$, *m-harmonic* everywhere except possibly on S. If u itself is m-harmonic, the volume integral in (6) vanishes. In general, by taking ∇^{2m} of both sides of (6), we see that $\nabla^{2m}u = \nabla^{2m}R_m$.

If P is outside S, as pointed out at the end of last section, the right hand member of (6) is equal to zero. Hence the forces of the surface and volume charges always balance out to a null-field outside of S.

It may be remarked that the problem of expanding an analytic function u in a series $\sum_1^\infty u_m$ such that u_m is m-harmonic admits of an infinite number of solutions: u_m, for instance, might be taken as the sum of the terms of degrees $2m - 2$, $2m - 1$ of a Taylor expansion for u about any point.

4. **Proof of (7) for Analytic Functions.** We shall now show that if u is analytic, (7) will hold if the region bounded by S is properly restricted; more precisely, we shall prove:

THEOREM. *If the Taylor series for u about a point $O \equiv (a_i)$ converges absolutely for $| x_i - a_i | = \rho_i$, then (7) will hold provided S lies within a sphere with center at O and of radius $\rho/3$, where ρ is the smallest of ρ_i; moreover, the integrand of (7) converges absolutely and the order of summation and integration may be interchanged.*

The point O, for convenience, will be taken at the origin. First we shall prove the lemma:

LEMMA. *If the Taylor series for u about O converges absolutely for $|x_i| \leqq \rho_i$, the series about $P \equiv (x_i)$ will converge at $P' \equiv (x_i + y_i)$ to u, the convergence will be absolute and the sum of the absolute values will be bounded for all P and P', provided that $| x_i | + | y_i | \leqq \rho_i$.*

Indeed, if the Taylor expansion of u about O, x_i is replaced by $x_i + y_i$ and the powers of the latter expanded and the terms rearranged in powers of y_i, there results the Taylor series about $P \equiv (x_i)$ evaluated at $P' \equiv (x_i + y_i)$. The rearrangement of terms is justifiable if $| x_i | + | y_i | \leqq \rho_i$. The convergence at P', moreover, is uniform, by Weierstrass's M-test if $| x_i | + | y_i | \leqq \rho_i$.

It follows from this lemma that if a sphere Σ be constructed with center at (x_i), $| x_i | < \rho_i$, tangent to the nearest of the planes $x_i = \pm\rho_i$, the Taylor series about the center (x_i) will converge absolutely at all the points of that spherical surface, the sum of the absolute values being less than or equal to a constant C. The radius of this sphere we shall denote by σ; it is equal to the smallest of $\rho_i \pm x_i$.

[4] That is, satisfying the equation $\Delta^{2i} = 0$.

If we write the Taylor series in the form

$$u(x_i + y_i) = \sum_{k=0}^{\infty} \frac{\partial^k u(x_i)}{\partial r^k} \frac{r^k}{k!}, \qquad r = (y_1^2 + \cdots + y_n^2)^{\frac{1}{2}},$$

then it will still be true that this series converges absolutely to a value less than C:

$$\sum_{k=0}^{\infty} | \partial^k u/\partial r^k | \sigma^k/k! \leqq C$$

for all points of Σ. Averaging this inequality[5] term by term over Σ, we get

$$\sum_{k=0}^{\infty} A(| \partial^k u/\partial r^k |) \sigma^k/k! \leqq C,$$

and hence

$$A(\partial^k u/\partial r^k)\sigma^k/k! \leqq A(| \partial^k u/\partial r^k |)\sigma^k/k! \leqq C.$$

Replacing k above by $2k$ and recalling

II (19) $A(\partial^{2k} u/\partial r^{2k}) = \nabla^{2k} u(2k)!/2\cdot 4 \cdots 2k\cdot n(n + 2) \cdots (n + 2k - 2),$

one obtains

$$(8) \qquad | \nabla^{2k} u |_{x_i} \leqq C\sigma^{-2k} 2\cdot 4 \cdots 2k\cdot n(n + 2) \cdots (n + 2k - 2).$$

The value of σ in (8) is equal to the smallest of $\rho_i \pm x_i$ and depends on x_i. If, however, S be restricted so that x_i are to be numerically less than $\rho_i/3$, then $\rho_i \pm x_i \geqq \rho_i - (\rho_i/3) = 2\rho_i/3 \geqq 2\rho/3$, where ρ is the smallest of ρ_i, and the inequality (8) will hold with σ replaced by $2\rho/3$, a value independent of x_i:

$$(8') \quad | \nabla^{2k} u |_{x_i} \leqq C\, 2\cdot 4 \cdots 2k\cdot n(n + 2) \cdots (n + 2k - 2)/(2\rho/3)^{-2k} \qquad \text{if } | x_i | \leqq \rho_i/3.$$

Now recall the remainder $R_m = \int V_m(r) \nabla^{2m} u\, dt$. If the surface S lies in a sphere with center at the origin and radius less than $\rho/3$, (8′) applies to the iterated Laplacian in the integrand. As regards the other factor, $V_m(r)$, if we consider the case of odd n first, it varies as r^{2m-n}, and, since r is the distance between two points inside S (if P is inside S), $| V_m(r) | < | V_m[\frac{2}{3}\rho(1 - \epsilon)] |$ for properly small ϵ, provided $m > n/2$. Hence, using the explicit form for V_m(II, §2, eqs. (35), (36)):

$$V_m(r) = r^{2m-n}/2\cdot 4 \cdots (2m - 2)[(2 - n)(4 - n) \cdots (2m - n)],$$

---

[5] By the *average* or *arithmetic mean*, $A(u)$, of a function u over Σ is understood, as usual, the quotient of the integral of the function over the surface divided by the surface: $\int u d\Sigma/\int d\Sigma$. The *averaging* operator A is used extensively in II. In the present case the mean A is taken with respect to the solid angle at $r = 0$.

and if $m > n/2$,

$$|R_m| < C\,\frac{2\cdot 4\cdots 2m\cdot n(n+2)\cdots(n+2m-2)}{(2\rho/3)^{2m}}$$

$$\cdot\frac{[2\rho(1-\epsilon)/3]^{2m-n}}{2\cdot 4\cdots(2m-2)\,|(2-n)\cdots(2m-n)|}$$

$$= C'\,\frac{(2n-1)(2n+1)\cdots(n+2m-2)\cdot 2m}{1\cdot 3\cdots(2m-n)}\,(1-\epsilon)^{2m},$$

where C, C' are constants independent of m. Using Stirling's formula, we may write

$$\frac{(2n-1)(2n+1)\cdots(n+2m-2)}{1\cdot 3\cdots(2m-n)} = \frac{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n}{2}+m\right)}{\Gamma\left(n-\frac{1}{2}\right)\Gamma\left(-\frac{n}{2}+m+1\right)}$$

in the form

$$\frac{\Gamma\left(\frac{1}{2}\right)}{\Gamma\left(n-\frac{1}{2}\right)}\,e^{-n+1}\,\frac{\left(\frac{n}{2}+m\right)^{\frac{n}{2}+m-\frac{1}{2}}}{\left(-\frac{n}{2}+m+1\right)^{-\frac{n}{2}+m+\frac{1}{2}}}\left\{1+O\left(\frac{1}{m}\right)\right\}$$

$$= \frac{\Gamma\left(\frac{1}{2}\right)}{\Gamma\left(n-\frac{1}{2}\right)}\,e^{-n+1}\,\frac{\left(1+\frac{n}{2m}\right)^{m}}{\left(1+\frac{1}{m}-\frac{n}{2m}\right)^{m}}\left(m^2+m+\frac{n}{2}-\frac{n^2}{4}\right)^{\frac{n}{2}-\frac{1}{2}}\left\{1+O\left(\frac{1}{m}\right)\right\}$$

$$= O\left(m^2+n^2+\frac{n}{2}-\frac{n^2}{4}\right)^{\frac{n}{2}-\frac{1}{2}} = O(m^{n-1}).$$

Therefore $|R_m| <$ Const. $m^n(1-\epsilon)^{2m}$ and R_m approaches zero as m becomes infinite. Formula (7) is thus proved for odd n under the conditions stated at the beginning of this section.

The case of even n is handled in a similar way. If $m > 1 + (n/2)$ (see II (35))

$$V_m(r) = \frac{r^{2m-n}\left[\log r - \left(\frac{1}{2}+\frac{1}{4}+\cdots+\frac{1}{2m-n}\right)+\left(\frac{1}{n}+\frac{1}{n+2}+\cdots+\frac{1}{2m-2}\right)\right]}{2\cdot 4\cdots(2m-2)\,[(2-n)\cdots(2m-n)]'}.$$

Break up V_m into an algebraic sum of two terms, the first one consisting of $r^{2m-n} \log r$ over the same denominator as above, and effect the corresponding division in R_m: $R_m = \int I_1 dt + \int I_2 dt$. Now,

$$| r^{2m-n} \log r | = | r \log r \cdot r^{2m-n-1} | < C[2\rho(1 - \epsilon)/3]^{2m-n}$$

if $m > 1 + (n/2)$ and S is restricted to lie in $| x_i | < \rho/3$. Hence,

$$\left| \int I_1 dt \right| < C'' \frac{(2n-1)(2n+1) \cdots (n+2m-2)2m}{2 \cdot 4 \cdots (2m-n)} (1-\epsilon)^{2m};$$

thus the first integral is seen to approach zero as m becomes infinite. As regards the second integral,

$$\frac{1}{2} + \cdots + \frac{1}{2m-n} + \frac{1}{n} + \cdots + \frac{1}{2m-2} < 2\left(\frac{1}{2} + \cdots + \frac{1}{2m}\right)$$

$$= 1 + \frac{1}{2} + \cdots + \frac{1}{m} < 1 + \log m;$$

hence, proceeding as above, we obtain

$$\int I_2\, dt = O[m^n(\log m + 1)(1-\epsilon)^{2m}] = o(1).$$

To complete the proof of the theorem of this section it remains to show that the order of summation and integration in (7) may be interchanged. Under the same conditions as above for the convergence of the Taylor series for $| x_i | \leqq \rho_i$, and for S lying in a sphere of radius $\rho/3$ we may establish the uniform convergence of the infinite series in (7). Indeed, from the above proof follows that, even for points of S

$$(9) \qquad \sum_{i > \frac{n}{2}} | \nabla^{2i} u V_i(r) | \leqq C \sum_{i > \frac{n}{2}} i^n (1-\epsilon)^i + C' \sum_{i > \frac{n}{2}} i^n \log i (1-\epsilon)^i;$$

thus $\sum \nabla^{2i} u V_i(r)$ is dominated by a convergent series of positive constants. Making use of the inequality

$$\left| \frac{\partial u}{\partial n} \right| = \left| \frac{\partial u}{\partial x_1} \cos(nx_1) + \cdots + \frac{\partial u}{\partial x_n} \cos(nx_n) \right| \leqq \left| \frac{\partial u}{\partial x_1} \right| + \cdots + \left| \frac{\partial u}{\partial x_n} \right|,$$

substituting for u in (9) the partial derivatives $\partial u/\partial x_1, \cdots, \partial u/\partial x_n$ and adding, we see that $\sum_{i=0}^{\infty} V_i \partial(\nabla^{2i} u)/\partial n$ converges absolutely and uniformly over the points of S. On the other hand,

$$\left| \frac{\partial V_i}{\partial n} \right| = \left| V_i'(r) \frac{\partial r}{\partial n} \right| \leqq | V_i'(r) |;$$

since $\lim_{i \to \infty} [V_i'(r)/V_{i-1}(r)] = 0$, (this is readily shown, from the explicit forms for V_i)

$$\left| \frac{dV_i(r)}{dr} \right| < C | V_{i-1}(r) |,$$

and hence

$$\sum \left| \frac{\partial V_i}{\partial n} \nabla^{2i} u \right| \leqq C \sum | V_{i-1}(r) \nabla^{2i-2}(\nabla^2 u) |, \qquad i > i_0 > 0.$$

Applying (9) with u replaced by $\nabla^2 u$ one proves that this part of the integrand of (7) also converges uniformly over the points of S.

5. **Extending the Range of Validity of** (7). The proof just given shows the sufficiency of certain conditions for the validity of (7) but gives no indication as to their necessity. One can show that they are actually necessary in some cases by letting $n = 1$ and choosing $u(x) = 1/(x - C)$, $C > 0$. The preceding work may be applied to the case when the number of dimensions is reduced to unity if familiar interpretations are made of the ideas of a region, its bounding surface S, the area of a unit sphere, etc., as an interval, its end points, 2, etc. The right hand member of (7) turns out to be half the sum of the Taylor expansions of the function at the end points of the interval (a, b) evaluated at a point P inside the interval (the interval (a, b) replaces the region t; the points a, b replace S). One readily shows that the series in question converges to u_P for all points in (a, b) where a, b are subject to an inequality $| a |, | b | < d$ when and *only* when $d = C/3$.

For special classes of functions, however, one may prove (7) under far less restrictive conditions. Thus, if u is k-harmonic, (7) holds for any S provided that u possesses continuous derivatives of order $2k$ within and on the boundary of S and satisfies there the equation $\nabla^{2k} u = 0$, since the remainder vanishes for $m > k$ and (7) thus reduces to the finite sum (6).

More generally, let u be a solution of

II (13): $$\sum_{i=0}^{p} c_i \nabla^{2i} u = 0;$$

it will be shown that (7) holds for any S within and on which the differential equation is satisfied.

Indeed, from II, §3 follows that the mean $A(u)$ of u over concentric spheres lying in a region in which u satisfies II (13) is an integral function of r^2, that is, expansible in a power series in r^2 convergent for all r. We may now identify this series with II (17) by applying the argument following II (54); to this end knowledge of the existence of $\nabla^{2k} u$ for any k is required. This existence is readily deduced from the differential equation by writing it in the form

$$\nabla^{2p} u = \sum_{i=0}^{p-1} (c_i/c_p) \nabla^{2i} u,$$

thus expressing ∇^{2p} in terms of lower order repeated Laplacians. It now follows that ∇^2 is applicable to the above equation; similarly it may be applied to the

resulting equation, and so forth. Now, if $\sum_{k=0}^{\infty} | C_k r^k |$ converges for $r = \sigma$ to the value C, $| C_k |$ satisfy the inequalities

$$| C_k | \leqq \frac{C}{[\sigma(1-\epsilon)]^k}.$$

Hence this inequality with an arbitrarily large σ will apply to the coefficients of the power series II (17):

II (17) $A(u) = u_0 + (\nabla^2 u)_0 r^2/2 \cdot n + (\nabla^4 u)_0 r^4/2 \cdot 4 \cdot n(n+2) + \cdots$

Now as shown in II §3, $A(u)$ is a linear combination of certain p integral functions of r^2 with coefficients which are linear in the values of $u, \nabla^2 u, \cdots \nabla^{2p-2} u$ at the center of the concentric spheres. Thus a single dominating series can be obtained for $A(u)$ for any position of the center O in the region R in which the differential equation is satisfied and hence the inequality obtained on the coefficients of II (17) applies with a constant C which depends only on σ but is independent of the position of O in R. The proof of the preceding section thus applied to all of R.

The differential equation enables us to eliminate $\nabla^{2p} u, \nabla^{2p+2} u, \cdots$ from the right hand member of (7) and obtain an expression for u in terms of $u, \nabla^2 u, \cdots, \nabla^{2p-2} u$ and their normal derivatives at the surface S. By applying ∇^{2k}, $k = 0, 1, 2, \cdots$, to the differential equation as above, there result what may be regarded as a system of linear algebraic recurrence equations in the values of $u, \nabla^2 u, \cdots$ at a point. This system obviously admits of a p-parameter family of solutions. If $\sum_{i=0}^{p} c_i x^i$ has simple roots, $\lambda_1, \cdots, \lambda_p$, the solutions are given by

$$\nabla^{2i} u = A_1 \lambda_1^i + A_2 \lambda_2^i + \cdots + A_p \lambda_p^i; \qquad i = 0, 1, 2, \cdots, \tag{10}$$

where the A's are (so far) arbitrary constants at each point. They may be solved for from the equations obtained by putting $i = 0, 1, \cdots, p-1$:

$$\begin{aligned} u &= A_1 + \cdots + A_p, \\ \nabla^2 u &= A_1\lambda_1 + \cdots + A_p \lambda_p, \\ &\cdots\cdots\cdots \\ \nabla^{2p-2} u &= A_1 \lambda_1^{p-1} + \cdots + A_p \lambda_p^{p-1} \end{aligned}$$

and expressed in terms of $u, \cdots, \nabla^{2p-2} u$. Substituting in (7) there results (since the integration may be carried out last)

$$\begin{aligned} u_P &= \int \sum_{k=1}^{p} \left[A_k \sum_{i=1}^{\infty} (\lambda_k)^{i-1} \frac{\partial V_i}{\partial n} - \frac{\partial A_k}{\partial n} \sum_{i=1}^{\infty} (\lambda_k)^{i-1} V_i \right] dS \\ &= \int \sum_{k=1}^{p} \left[A_k \frac{\partial \psi(r, \lambda_k)}{\partial n} - \frac{\partial A_k}{\partial n} \psi(r, \lambda_k) \right] dS, \end{aligned} \tag{11}$$

where $\psi(r, \lambda) = \sum_{i=1}^{\infty} V_{i+1}(r) \lambda^i$ and is a solution of $\nabla^2 u - \lambda u = 0$; and, eliminating A_k,

$$
u_p \begin{vmatrix} 1 & 1 & \cdot\cdot & 1 \\ \lambda_1 & \lambda_2 & \cdot\cdot & \lambda_p \\ \cdot & \cdot & \cdot\cdot & \cdot \\ \cdot & \cdot & \cdot\cdot & \cdot \\ (\lambda_1)^{p-1} & (\lambda_2)^{p-1} & \cdot\cdot & (\lambda_p)^{p-1} \end{vmatrix} = \int - \begin{vmatrix} 0 & \dfrac{\partial\psi(r, \lambda_1)}{\partial n} & \cdot\cdot & \dfrac{\partial\psi(r, \lambda_p)}{\partial n} \\ u & 1 & \cdot\cdot & 1 \\ \nabla^2 u & \lambda_1 & \cdot\cdot & \lambda_p \\ \cdot & \cdot & \cdot\cdot & \\ \cdot & \cdot & \cdot\cdot & \\ \nabla^{2p-1} u & (\lambda_1)^{p-1} & \cdot\cdot & (\lambda_p)^{p-1} \end{vmatrix}
$$

$$
(12) \qquad + \begin{vmatrix} 0 & \psi(r, \lambda_1) & \cdot\cdot & \psi(r, \lambda_p) \\ \dfrac{\partial u}{\partial n} & 1 & \cdot\cdot & 1 \\ \dfrac{\partial(\nabla^2 u)}{\partial n} & \lambda_1 & \cdot\cdot & \lambda_p \\ \cdot & \cdot & \cdot\cdot & \cdot \\ \cdot & \cdot & \cdot\cdot & \cdot \\ \dfrac{\partial(\nabla^{2p-2} u)}{\partial n} & (\lambda_1)^{p-1} & \cdot\cdot & (\lambda_p)^{p-1} \end{vmatrix} dS.
$$

If $\sum_{i=0}^{p} c_i x^i$ has repeated roots, thus, if λ_1 is a k_1-fold root, λ_2 a k_2-fold root, $\cdots$, λ_s a k_s-fold root, $(k_1 + k_2 + \cdots + k_s = p)$, the solutions of the above system of linear algebraic equations are given by[6]

$$
\begin{aligned}
\nabla^{2i} u = {} & A_1(\lambda_1)^i + A_2 \frac{d\lambda_1^i}{d\lambda} + \cdots + A_{K_1} \frac{d^{K_1-1}(\lambda_1^i)}{d\lambda^{K_1-1}} \\
& + A_{K_1+1}(\lambda_2)^i + \cdots + A_{K_1+K_2} \frac{d^{K_2-1}(\lambda_2^i)}{d\lambda^{K_2-1}} \\
& \qquad \cdot \qquad \cdots \qquad \cdot\cdot \\
& + A_{p-K_s+1}(\lambda_s)^i + \cdots + A_p \frac{d^{K_s-1}(\lambda_s^{K_s})}{d\lambda^{K_s-1}}.
\end{aligned}
$$

Hence we obtain an equation for u_P that differs from (12) by having certain columns in the determinants replaced by their λ-derivatives. The determinant that multiplies u_P does not vanish.

Equation (12) and its indicated modifications could also be established without the use of (7) as follows. Consider the case of simple roots; $A_1, A_2, \cdots, A_p$ whose sum is equal to u satisfy respectively the differential equations $(\nabla^2 - \lambda_i)A_i = 0$. By applying to each one of them Green's theorem in the form

$$
\int[u(\nabla^2 v - \lambda_i v) - v(\nabla^2 u - \lambda_i u)]\, dt = \int(u\, \partial v/\partial n - v\, \partial u/\partial n)\, dS,
$$

choosing $u = A_i$, $v = \psi(r, \lambda_i)$, one obtains, as in §2,

$$
K_n(A_i)_P = \int[A_i\, \partial\psi(r, \lambda_i)/\partial n - \psi(r, \lambda_i)\, \partial A_i/\partial n]\, dS,
$$

[6] This representation leads to the inequality $|\nabla^{2k} u| <$ Const. $k(k-1) \cdots (k-p+1)\Lambda^k$, where Λ is the largest absolute value of the roots λ_i. This inequality is an improvement on (8).

and, eliminating A_i from the sum of these integrals, one is led to (12). In case of multiple roots, however, the direct proof along these lines is complicated. An alternative method that would apply in all cases consists in properly modifying the scheme used in §2 by replacing in equations (1) ∇^2 by $(\nabla^2 - \lambda_i)$ and carrying out similar modifications in (2) and (3).

If all the p roots are equal to zero there results the multiharmonic case which was mentioned at the beginning of this section; (12) or rather, its indicated modification, reduces to

$$(13) \quad K_n u_P = \int [u\, \partial V_1/\partial n - V_1\, \partial u/\partial n] + [\nabla^2 u\, \partial V_2/\partial n - V_2\, \partial(\nabla^2 u)/\partial n] + \cdots + [\nabla^{2p-2} u\, \partial V_p/\partial n - V_p\, \partial(\nabla^{2p-2} u)/\partial n]\, dS.$$

More generally, if all the roots are equal to λ, one obtains

$$K_n u_p = \int \Bigg[u \frac{\partial \psi(r, \lambda)}{\partial n} - \psi(r, \lambda) \frac{\partial u}{\partial n} \Bigg] + \frac{1}{1!} \Bigg[(\nabla^2 u - \lambda u) \frac{\partial}{\partial n} \frac{\partial \psi(r, \lambda)}{\partial \lambda} - \frac{\partial \psi(r, \lambda)}{\partial \lambda} \frac{\partial}{\partial n} (\nabla^2 n - \lambda u) \Bigg] + \cdots + \frac{1}{(p-1)!} \Bigg[(\nabla^2 - \lambda)^{p-1} u \frac{\partial}{\partial n} \frac{\partial^{p-1} \psi(r, \lambda)}{\partial \lambda^{p-1}} - \frac{\partial^{p-1} \psi(r, \lambda)}{\partial \lambda^{p-1}} \frac{\partial}{\partial n} [(\nabla^2 - \lambda)^{p-1} u] \Bigg] dS.$$

The integrals in the brackets above satisfy respectively the differential equations $(\nabla^2 - \lambda) = 0$, $(\nabla^2 - \lambda)^2 = 0, \cdots$.

The last result at once suggests an infinite series for analytic functions analogous to (7) with $\nabla^2 - \lambda$ taking the place of the Laplacian and V_i replaced by $\partial^{i-1}\psi(r, \lambda)/\partial\lambda^{i-1}(i - 1)!$. Without at present attempting a proof of such expansions along the lines of §2 and §4, we point out that the integrand in the resulting expansion

$$(15) \quad u_P = \int \sum_{i=0}^{\infty} \Bigg[\frac{(\nabla^2 - \lambda)^i u}{i!} \frac{\partial}{\partial n} \frac{\partial^i \psi(r, \lambda)}{\partial \lambda^i} - \frac{\partial}{\partial n} \frac{(\nabla^2 - \lambda)^i u}{i!} \frac{\partial^i \psi(r, \lambda)}{\partial \lambda^i} \Bigg] dS$$

is equal to the integrand of (7) provided that certain rearranging of terms is permissible.

By Taylor's expansion, if x is any number,

$$\sum_{i=0}^{\infty} \frac{(x - \lambda)^i}{i!} \frac{\partial^i \psi(r, \lambda)}{\partial \lambda^i} = \sum_{i=0}^{\infty} \frac{x^i}{i!} \frac{\partial^i \psi(r, \lambda - \lambda)}{\partial \lambda^i} = \sum_{i=0}^{\infty} x^i V_{i+1}(r).$$

From this identity in x follows the (formal) identity in u

$$\sum_{i=0}^{\infty} \frac{\partial}{\partial n} \frac{(\nabla^2 - \lambda)^i u}{i!} \frac{\partial^i \psi(r, \lambda)}{\partial \lambda^i} = \sum_{i=0}^{\infty} \frac{\partial}{\partial n} \nabla^{2i} u\, V_{i+1}(r)$$

by grouping together terms with the same $\partial(\nabla^{2i} u)/\partial n$; the coefficients of $\partial(\nabla^{2i} u)/\partial n$ are the same as the coefficients of x^i in the above identity. If,

again, similar grouping of terms be permissible in the former part of the integrand of (11), it will be equal to the sum $\Sigma\nabla^{2i-2}u\ \partial V_i/\partial n$ of the integrand of (7).

Further generalizations of such type may be attempted by replacing ∇^2 in the equations (1) by $(\nabla^2 - \lambda_i)$ with λ_i varying from one equation to the other. It is intended to treat some of these generalizations at a future date.

6. **Examples.** We shall now give a few applications of the preceding formulas.

A. Choose $u = V_k(r')$, where r' is the distance from a fixed point P'. If P' is outside S while P is inside we obtain from (6) (since $\nabla^2 u = V_{k-1}(r')$, $\nabla^4 u = V_{k-2}(r'), \cdots$)

$$u_P = V_k(PP') = \frac{1}{K_n}\sum_{i=1}^{P}\left[V_{k-i}(r')\frac{\partial V_i(r)}{\partial n} - \frac{\partial V_{k-i}(r')}{\partial n}V_i(r)\right]dS. \tag{16}$$

If P and P' are both outside S, the surface integral above vanishes. Such is also the case when both points are inside S. For, returning to the proof of (6) in Section 2, we see that both v_i and u_i possess singularities inside S (at P and P', respectively); consequently, the neighborhoods of both points have to be excluded in applying Green's theorem, say, by small spheres Σ, Σ' with centers at P and P'. If the radii of the latter be made equal, the integrals over Σ will cancel in reverse order the integrals over Σ'. Thus the above surface integral is equal to $V_k(PP')$ or to zero depending upon whether or not the two points are separated by S. Should one of the points fall on S the integral will give (under proper restrictions on S) $V_k(PP')/2$; finally, if both points fall on S the value of the integral is, again, equal to zero.

Similar integrals could be obtained from (14). Thus, putting $u = \psi(r', \lambda)$, where r' has the same meaning as above, we obtain

$$\frac{1}{K_n}\int\left[\psi(r', \lambda)\frac{\partial\psi(r, \lambda)}{\partial r} - \frac{\partial\psi(r', \lambda)}{\partial n}\psi(r, \lambda)\right]dS = \begin{Bmatrix}\psi(PP')\\ \psi(PP')/2\\ 0\end{Bmatrix} \text{ as above.} \tag{17}$$

B. As an application of (13), (14) we consider the evaluation of the spherical means of functions such as $V_p(r)$, $\psi(r, \lambda)$, $\cdots$ over spheres enclosing the singular point P: $r = 0$. We shall indicate the derivation of the following equations:

$$\begin{aligned}\int V_p(r_1)\,dS/S = V_p(r) + V_{p-1}(r)\frac{a^2}{2\cdot n}\\ + \cdots + \frac{V_1(r)a^{2p-2}}{2\cdots(2p-2)n\cdots(n+2p-4)},\end{aligned} \tag{18}$$

$$\int\frac{\partial^{p-1}\psi(r_1, \lambda)}{\partial\lambda^{p-1}}dS/S = \frac{\partial^{p-1}\psi(r, \lambda)}{\partial\lambda^{p-1}}\phi(a, \lambda) + \cdots + \psi(r, \lambda)\frac{\partial^{p-1}\phi(a, \lambda)}{\partial\lambda^{p-1}} \tag{19}$$

for $r > a$. Here the left hand integration is over a sphere S of radius r and center O; S is also used for the area $K_n r^{n-1}$ of this sphere; r_1 is the distance from a fixed point P_1 inside S; $a = OP_1$ is the radius of the sphere concentric

with S and passing through P_1, while $\phi(r, \lambda)$ is the symmetric solution of II (1) which is analytic at $r = 0$ (see I, Sections 1, 2). The integrands in (18), (19) are solutions of II (11), II (8) respectively except at P_1, and application of II (12), II (9) toward the evaluation of their spherical means over spheres S *not* enclosing P_1 shows that for $r < a$, a and r should be interchanged on the right-hand sides of (18), (19). The treatment of the case $r > a$ by the method briefly indicated at the end of II (II, concluding paragraph section 9) and based on *spreading* the singularity at P_1 will be given in a forthcoming paper "On Integral Representations of Bessel and Related Functions." We proceed with the present proof.

First consider (19) for the case $p = 1$:

$$\int \psi(r_1, \lambda)\, dS/S = \psi(r, \lambda)\phi(a, \lambda). \tag{19'}$$

Expanding both sides in powers of λ leads to (18) when the coefficients of λ^p on both sides are equated.

Apply (14) for $p = 1$ ever the sphere S of radius r_0 to the function $\phi(r, \lambda)$: There results

$$K_n\phi(r, \lambda) = \phi(r_0, \lambda) \int \frac{\partial\psi(r_1, \lambda)}{\partial n}\, dS - \frac{\partial\phi(r, \lambda)}{\partial r}\bigg|_{r=r_0} \int \psi(r_1, \lambda)\, dS. \tag{20}$$

Next apply (17) with P' at O, over the same sphere (replacing r' by r_1). Since P and P' are both inside S, there results

$$0 = \psi(r_0,\lambda) \frac{\partial\psi(r_1, \lambda)}{\partial n}\, dS - \frac{\partial\psi(r, \lambda)}{\partial r}\bigg|_{r=r_0} \int \psi(r_1, \lambda)\, dS. \tag{20'}$$

By solving (20) (20′) for $\int \psi(r_1, \lambda)\, dS$ and replacing r_0 by r one obtains (19′), provided the determinant of the coefficients of the two integrals be simplified by means of

$$\begin{vmatrix} \phi(r, \lambda) & \partial\phi(r, \lambda)/\partial\lambda \\ \psi(r, \lambda) & \partial\psi(r, \lambda)/\partial\lambda \end{vmatrix} = r^{1-n}.$$

To obtain the latter allow r to approach O in (20); the integrands become constant over S and $\int dS$ may be replaced by multiplication by $S = K_n r_0^{n-1}$.[7]

Another way of viewing the above derivation is by applying (14) with $p = 1$ to the symmetric solution of $(\nabla^2 - \lambda)u = 0$ which vanishes along S. Similarly (19) may be derived for any p by applying (14) to that symmetric solution of $(\nabla^2 - \lambda)^p u = 0$ for which $u = \partial u/\partial r = \nabla^2 u = \cdots = \nabla^{2p-2}u = 0$ at $r = r_0$; of course, the singularity of u at $r = 0$ has to be attended to.

The integral $\int \partial\psi(r_1, \lambda)/\partial n\, dS$ whose value may be found from (20), (20′)

---

[7] This reduction is also apparent from the fact that the determinant is the Wronskian of the two solutions ϕ, ψ of II (1):

$$(\nabla^2 - \lambda)u = [d^2/dr^2 + (n - 1)\, d/dr - \lambda]u = 0.$$

may also be found by differentiating (19′) with respect to the radius of S; it is convenient to replace the surface integrations and means by integrations with respect to the solid angle $d\omega$ at the center before differentiation. Similarly, differentiation of (18), (19) with respect to r evaluate $\int \frac{\partial V_p(r_1)}{\partial n} d\omega / (\int d\omega = K_n)$, etc.

C. Let $u(r)$ be an even analytic function of the complex variable r, nonsingular at the origin $r = 0$. Interpret r as the distance from a fixed point in a space on n dimensions, and apply (7) to $u(r)$ choosing for the surface S the sphere $r = a$. If P is now any point inside this sphere at a distance r from the center and r_1 is the distance from P, there results the series

$$(21)\quad u(r) = \frac{1}{K_n} \sum_{i=1}^{\infty} \left[(\nabla^{2i-2} u)_{r=a} \int \frac{\partial V_i(r_1)}{\partial n} dS - \frac{d}{dr} (\nabla^{2i-2} u)_{r=a} \int V_i(r_1)\, dS \right]$$

where ∇^{2k} are to be replaced by $[d^2/dr^2 + \{(n-1)/r\}\, d/dr]^k$. The convergence of this series to $u(r)$ for small enough r follows at once from the validity of (7). The integrals in (21) are precisely the integrals considered in (18) (but with a, r interchanged) and their derivatives. The expansion is thus of the form

$$(21')\qquad u(r) = \sum_{i=0}^{\infty} \nabla^{2i} u \Big|_{r=a} \frac{\partial P_i(r, a)}{\partial r} - \frac{d(\nabla^{2i} u)}{dr} \Big|_{r=a} P_i(r, a)$$

where P_i are polynomials in r^2. These polynomials may be proved to satisfy the recurrence relations

$$(22)\qquad \nabla^2 P_i(r) = \left(\frac{d^2}{dr^2} + \frac{n-1}{r} \frac{d}{dr} \right) P_i(r) = P_{i-1}(r).$$

For $n = 1$ the expansion reduces to the Taylor series about a evaluated at r.

It is of interest to note that by applying (15) in a similar manner to an even analytic function of r one is led to an expansion in terms of the functions appearing in (19); these functions, it will be recalled, are expressible in terms of Bessel functions. It is intended to treat these expansions elsewhere.

D. As another application we shall indicate the evaluation over sphere S of integrals

$$(23)\qquad \int_S \frac{\partial^{p-1} \psi(r, \lambda)}{\partial \lambda^{p-1}} h_k(\omega')\, d\omega', \cdots,$$

where $h_k(\omega')$ is a surface spherical harmonic over S of degree k, $dS = r^{n-1} d\omega'$ being the element of area of S. These integrals generalize (18), (19) and are of interest in connection with expansions (to be considered elsewhere) resembling the Neumann expansions. Again the case where r_1 is measured from a point outside S can be treated by the methods indicated in II (§9, modification of (98) to the case of equal λ).

To evaluate (23) we proceed as in case of (19) by applying (14) to a proper function u. The latter is now chosen to be of the form

$$u = h_k(\omega') I(r)$$

and so as to satisfy $(\nabla^2 - \lambda)^p = 0$. As shown in II, §9, $I(r)$ satisfies the differential equation II (93) and consists of linear combinations of the functions II (94); this combination is picked again so that $u = \partial u/\partial r = \nabla^2 u = \cdots = \nabla^{2p-2} = 0$ along S. The singularity at the center requires special treatment, and an application of II (89) is found very useful in evaluating the limits of the integrals over a sphere excluding the center.

7. **Analyticity of Solutions of $\sum_{i=0}^{p} c_i \nabla^{2i} u = 0$ and Their Taylor Expansions.** We shall now utilize the expressions of the solutions of the differential equation II (13) in terms of boundary values given in §5 to show the analyticity of such solutions and examine the range of convergence of their Taylor expansions.

Let the differential equation be satisfied within and on a sphere S of radius a and center O. Let $\Pi \equiv (\zeta_i)$ be any point on S, $P \equiv (x_i)$ any point within S and let

$$r = OP, \qquad r' = \Pi P, \qquad \theta = \text{angle } \Pi OP, \qquad \delta = \text{angle } P\Pi O.$$

Consider $r'^{2k} = [a^2 + r^2 - 2ar\cos\theta]^k = a^{2k}[(1 - re^{i\theta}/a)(1 - re^{-i\theta}/a)]^k$ where k is any (real) constant. Since $(1 - z)^k$ can be expanded as a power series in z that converges for $|z| < 1 - \epsilon$, $\epsilon > 0$ uniformly and absolutely, we may expand $(1 - re^i\theta/a)^k$ as well as $(1 - re^{-i\theta}/a)^k$ in powers of r/a; the two series may be multiplied and the product arranged in powers of r, and the resulting series will converge absolutely and uniformly to r'^{2k} for θ real and for $|r| < a - \epsilon'$. A similar conclusion applies to the expansion of $\log r'$ in powers of r. Now $\psi(r', \lambda) = r'^{2-n} P_1(r'^2) + \log r' \, P_2(r'^2)$, where P_1, P_2 are certain power series representing entire functions of their argument. If in these power series we replace r'^2 by $[a^2 + r^2 - 2ar\cos\theta]$ and expand the powers of the bracket in powers of r, the resulting series is dominated by the series of absolute values of the terms that result when $a^2 + r^2 + 2ar$ is substituted for r'^2. We may therefore expand $P_1(r'^2)$, $P_2(r'^2)$ in powers of r and the resulting series will converge absolutely and uniformly for $|r| < a - \epsilon'$ and θ real. From this it follows that $\psi(r', \lambda)$ may be expanded in powers of r, the series converging uniformly for all the triangles $OP\Pi$, where Π is on S and P within a concentric S_1 of radius $a - \epsilon'$. Similar expansions, it can be shown, hold for $\partial\psi(r'\lambda)/\partial\lambda$ which may occur in the integrand of (12). This also involves the functions $\partial\psi(r', \lambda)/\partial n = (\partial r'/\partial n)\, \partial\psi(r', \lambda)/\partial r'$

$$= \frac{\partial\psi(r'\lambda)}{\partial r'}\cos\delta = \frac{\partial\psi(r', \lambda)}{\partial r'}\,\frac{a - r\cos\theta}{r}$$

$$= [r'^{-n} P_1(r'^2) + \log r' \, P_2(r'^2)]\,(a - r\cos\theta),$$

where P_1, P_2, as above, represent power series in the argument r'^2 (though not the same series as before) converging for all values of the argument. These

parts of the integrand may be expanded in powers of r, and the series will converge uniformly for the same range as above. We may thus expand the integrand of (12) in such a series and integrate term by term over S, since the integrand consists of a bilinear combination with constant coefficients of $\psi(r, \lambda)$, $\partial\psi(r, \lambda)/\partial n, \cdots$ and continuous functions over S. The result will converge uniformly to u_P at all P within (and on) a sphere S_1 of radius less than a.

The expansion of r'^{2k} in powers of r may also be carried out by applying the binomial theorem to

$$[a^2 + (r^2 - 2ar \cos \theta)]^k$$

and arranging terms. This is permissible for small enough r but the resulting series must be identical (for each fixed θ) with the series formerly obtained, since an element of an analytic function admits of only one power series expansion. We thus see that any term in the expansion of r'^{2k} in powers of r is given by a linear combination with constant coefficients of products of r^2 and $r \cos \theta$, and is therefore a homogeneous polynomial (for a fixed Π) in x_i, the coordinates of P. The same thing is true of the terms of the expansion of

$$\psi(r', \lambda),\ \partial\psi(r', \lambda)/\partial n,\ \partial\psi(r', \lambda)/\partial\lambda,\ \cdots$$

in powers of r. Hence the terms of the expansion found for u are homogeneous polynomials in x_i. Solutions of II (13) are thus expansible in series of homogeneous polynomials that converge in the largest sphere in which they possess continuous derivatives of order $2p$ and satisfy the differential equation. A special case of these is the expansion of a harmonic function in spherical harmonics.

Turning to the question of a Taylor series for u consider the expansion in powers of x_i of

$$\begin{aligned} r'^{2k} &= [(x_1 - \zeta_1)^2 + (x_2 - \zeta_2^2) + \cdots + (x_n^2 - \zeta_n)^2]^k \\ &= [a^2 - 2(x_1\zeta_1 + \cdots + x_n\zeta_n) + (x_1^2 + x_2^2 + \cdots + x_n^2)]^k. \end{aligned}$$

Expand this in powers of $[2(x_1\zeta_1 + \cdots + x_n\zeta_n) - (x_1^2 + \cdots + x_n^2)]$, further expand these powers and collect terms in a multiple series in x_i. This series is dominated by the similar expansion for

$$[a^2 - 2(|x_1\zeta_1| + |x_2\zeta_2| + \cdots + |x_n\zeta_n|) - (x_1^2 + \cdots + x_n^2)]^{-|k|}$$

a series of positive terms that converges if

$$2(|x_1\zeta_1| + \cdots + |x_n\zeta_n|) + (x_1^2 + \cdots + x_n^2)^2 < a^2,$$

that is, if

$$(|x_1| + |\zeta_1|)^2 + \cdots + (|x_n| + |\zeta_n|)^2 < 2a^2.$$

The left hand member of this inequality represents the square of the distance from $| x_i |$ to $| \zeta_i |$ and cannot exceed $(a + r)^2$. The inequality will thus be satisfied if

$$a + r < \sqrt{2}a,$$

or if

$$r < (\sqrt{2} - 1)a = \rho.$$

If P lies within a sphere of radius ρ the convergence of the power series for r'^{2k} is uniform in P and Π. We can now complete the proof that u can be expanded in a Taylor series that converges uniformly and absolutely within a sphere of proper size. First one would show that the Taylor series for $\log r$ converges uniformly and absolutely in P and Π for $r \leqq \rho - \epsilon$; then, utilizing the everywhere absolutely convergent Taylor series for $P_1(r'^2)$, $P_2(r'^2)$, one could prove that the integrand of (12) may be expanded in such a series, which may be integrated term by term. The solutions of the differential equation II (13) are thus analytic in a region in which they satisfy the equation.

The former expansions in homogeneous polynomials are the same as the Taylor expansions if the latter are summed in groups of homogeneous terms. The terms of any one degree will be k-harmonic if u is k-harmonic; except for the p-harmonic case, however, these polynomials will not satisfy the differential equation satisfied by u.

By considering the expansions of r', $(r')^k$, $\psi(r', \lambda,) \cdots$ in positive powers of a (rather than of r), one obtains expansions proceeding in decreasing powers of r and valid for $| r | > | a |$. This follows from the fact that r' is symmetric in r and a. From these one can obtain expansions of certain solutions of II (13) in regions outside a spherical surface, and by combining with the preceding expansions, of solutions in spherical shells.

General Electric Company,
Schenectady, New York.

ANNALS OF MATHEMATICS
Vol. 39, No. 1, January, 1938

THE ASYMPTOTIC WARING PROBLEM FOR HOMOGENEOUS POLYNOMIAL SUMMANDS

BY H. CHATLAND

(Received June 19, 1937; revised September 14, 1937)

1. **Introduction.** In this paper, by using homogeneous polynomials as summands, the number required to represent all sufficiently large integers is reduced to approximately two-thirds of that for n^{th} powers. The asymptotic Waring Problem for sums of n^{th} powers given by I. Vinogradow[1] was generalized and the result improved by L. E. Dickson,[2] who considered sums of the type

$$Ax^n + \sum_{j=1}^{s-3} A_j x_j^n + \sum_{i=1}^{k} a_i(w_i^n + W_i^n) + \sum_{i=1}^{k} a_{k+i} z_i^n y^n, \tag{1}$$

where the coefficients are given positive integers.

If $s \geqq 4n$, the congruence

$$Ax^n + \sum_{j=1}^{s-3} A_j x_j^n + a_k(w^n + W^n) \equiv G (\operatorname{mod} p^\gamma) \tag{2}$$

is known to have integral solutions such that not every term on the left is divisible by p, for every integer G and prime p and γ defined to be $\theta + 1$ if $p > 2$, $\theta + 2$ if $p = 2$, p^θ denoting the highest power of the prime p which divides n.

In this paper we treat the problem for sums of values of a homogeneous polynomial

$$Ax^n + \sum_{j=1}^{s-3} A_j x_j^n + \sum_{i=1}^{k+1} \{\phi(x_i, y_i) + \phi(X_i, Y_i)\} + \sum_{i=1}^{k+1} y^n \phi(z_i, w_i), \tag{3}$$

where the coefficients are given positive integers and

$$\phi(x, y) = \sum_{r=0}^{n} b_r x^{n-r} y^r, \tag{4}$$
$$b_0 = b_1 = b_2 = 1, \qquad 0 \leqq b_i \leqq (n+1)^{i-2}(n-1)^{i-3} \qquad (i \geqq 3).$$

The congruence corresponding to (2) is

$$Ax^n + \sum_{j=1}^{s-3} A_j x_j^n + (\phi(x_{k+1}, y_{k+1}) + \phi(X_{k+1}, Y_{k+1})) \equiv G (\operatorname{mod} p^\gamma), \tag{5}$$

[1] I. Vinogradow, *On Waring's problem*, Annals of Mathematics, vol. 36 (1935), pp. 395–405.

[2] L. E. Dickson, *On Waring's problem and its generalization*, Annals of Mathematics, vol. 37 (1936), pp. 293–316.

which reduces to (2) when y_{k+1}, Y_{k+1} are zero. Hence (5) has solutions of the type described for (2) when $s \geqq 4n$. Apart from constants, the argument is the same and the result holds when the coefficients a_i are put into the third and fourth sums of (3). We obtain the

THEOREM. *Every sufficiently large integer is a sum of $4n - 2 + 3k_0$ values of a homogeneous polynomial of the type described, where k_0 is the least integer k satisfying*

$$k > \frac{\log r}{-\log (1 - 3\nu/2)},$$

with $\nu = 1/n$, $r = n^2(6n - 7 + \nu)/(n - d)$, $d = 1 + 2n^2 z$ and z is about $\nu^3/24$, $n \geqq 4$.

The notations given by Vinogradow and Dickson are used in this paper. After several new steps in the theory we are able to use a part of Dickson's paper, with his notations having the new meaning given them for this problem.

Section 4 has no analogue in the problem for n^{th} powers and notations used there pertain only to that part.

2. Notations and definitions.

Write

$$(6) \qquad \nu = 1/n, \qquad \sigma = (n-1)(1 - 3\nu/2)^k.$$

Let c_i, h, and f denote positive numbers depending only upon n. Given a large integer N_0, we take

$$(7) \qquad P = [hN_0^\nu], \qquad R = [P^{1-f}], \qquad m = \max(3^{n-1}A, A_1, \cdots, A_{s-3}),$$

$$(8) \qquad P_1 = [(n/2^{n+1})^\nu P^{1-\nu}], \qquad R_1 = \left[\left(\frac{n}{2^{n+1} - 1}\right)^\nu R^{1-\nu}\right], \qquad \tau = 2mnP^{n-1/2},$$

$$(9) \qquad Y = [C^\nu P^{(1-\nu)f}], \qquad C = (3/2)^{n-1}.$$

Then

$$(10) \qquad \tau \leqq P^n \text{ if } P \geqq (2mn)^2, \qquad Y^n \leqq \tfrac{1}{2}P^{\frac{1}{2}} \text{ if } P^{\frac{1}{2}-(n-1)f} \geqq 2C.$$

The latter will be true when N_0 (and hence P) is sufficiently large and

$$(n-1)f < \tfrac{1}{2}.$$

When z is real, $(z]$ denotes the distance from z to the nearest integer. When x is complicated, we denote e^x by Exp x. The largest integer $\leqq x$ is denoted by $[x]$.

3. Five lemmas.[3]

LEMMA A. *If a, q, t are positive integers, $(a, q) = 1$, then*

$$\left|\sum_{r=0}^{q-1} \text{Exp } 2\pi i \frac{a}{q} tr^n\right| < cq^{1-\nu}, \qquad c(n, t) > 0.$$

[3] L. E. Dickson, *ibid.*

LEMMA B. *When $s \geqq 4n$, $t_j = 1$, $n \geqq 4$, there exist positive constants $b = b(N, n, s)$, $c_5 = c_5(n, s)$, $c_2 = c_2(n, s)$, such that*

$$\left| \sum_{q=1}^{u} A(q, N) - b \right| \leqq c_5 u^{2-s\nu}, \qquad b > c_2, \tag{11}$$

where for positive integers q, s, t_j,

$$A(q, N) = \sum \operatorname{Exp}\left(-2\pi i \frac{a}{q} N\right) \prod_{j=1}^{s} \frac{1}{q} \sum_{r=0}^{q-1} \operatorname{Exp} 2\pi i \frac{a}{q} t_j r^n.$$

LEMMA B′. *For every prime p and every integer G, let*

$$\sum_{j=1}^{s} t_j r_j^n \equiv G \ (mod\ p^\gamma)$$

have a special solution. Let $s > 2n$, $s \geqq 5$. *Let t be the maximum of $t_1, \cdots, t_s$. Then there exist positive constants $b = b(N, n, s, t)$, $c_i = c_i(n, s, t)$, satisfying* (11).

LEMMA C. *Let $0 \leqq f'(y) \leqq \frac{1}{2}$, $f''(y) \geqq 0$ in the interval $g \leqq y \leqq h$, where g and h need not be integers. Then*

$$\left| \sum_{g<y\leqq h} \operatorname{Exp} 2\pi i f(y) - \int_g^h \operatorname{Exp} 2\pi i f(y)\, dy \right| < 5.$$

LEMMA D. *Let λ be real, but not integral. Let G and H be integers, $G < H$. Then*

$$\left| \sum_{x=G+1}^{H} \operatorname{Exp} 2\pi i \lambda x \right| < 1/2(\lambda].$$

4. **Some inequalities.** We shall determine a rectangle in the xy-plane which is such that when the coördinates of certain points in it are substituted in the polynomial

$$\phi(x, y) = \sum_{r=0}^{n} b_r x^{n-r} y^r, \tag{12}$$

with $b_0 = b_1 = b_2 = 1$, $0 \leqq b_i \leqq (n+1)^{i-2}(n-1)^{i-3}$ $(3 \leqq i \leqq n)$, the values of the polynomial will be distinct and the minimum difference between any two of them will be greater than a certain fixed quantity depending upon n and the least value P_1 of x in the rectangle considered. The following inequalities are necessary.

$$\begin{aligned} &\phi(P_1, A(n-1)) > \phi(P_1 + n - 1, (A - n)(n-1)) \\ &\qquad\qquad (n \leqq A \leqq P_1/(n+1)(n-1)^2), \end{aligned} \tag{I}$$

in which A is a multiple of n. Compare i^{th} terms of the polynomials (I). We find that

$$b_i(n-1)^i A^i P_1^{n-i} > b_i(n-1)^i (A-n)^i (P_1 + n - 1)^{n-i}. \tag{13}$$

This written as a difference is

$$(14)\qquad b_i(n-1)^i\{A^iP_1^{n-i}-(A-n)^i(P_1+n-1)^{n-i}\}>0,$$

which is true if $\left(\frac{A}{A-n}\right)^i P_1^{n-i} > (P_1+n-1)^{n-i}$. Now $A/(A-n)$ is minimum for A maximum and hence for $A = P_1/(n+1)(n-1)^2$. Then

$$A/(A-n) = 1 + n/(A-n) > 1 + n/A = 1 + n(n+1)(n-1)^2/P_1.$$

Thus (14) holds if

$$(1+n(n+1)(n-1)^2/P_1)^iP_1^{n-i} > P_1^{n-i} + (n-i)(n-1)P_1^{n-i-1} + (\text{terms in lower powers of } P_1).$$

This is true since on expanding the left side and dropping P_1^{n-i} we find that the second term on the left is greater than the remaining part on the right. That is

$$in(n+1)(n-1)^2P_1^{n-i-1} > (n-i)(n-1)P_1^{n-i-1} + \text{the remaining terms}.$$

Hence in (I) set $b_i = 0$ $(i \geqq 3)$. Subtracting the right side from the left we find the difference to be greater than

$$(A(n+1)-n(n+1)/2)\,P_1^{n-2} - (\text{terms in lower powers of } P_1).$$

Hence for P_1 large the difference between the right and left sides of (I) exceeds

$$(15)\qquad A\binom{n-1}{2}P_1^{n-2}$$

for $n \leqq A \leqq P_1/(n+1)(n-1)^2$. When $A = Q$, where Q is the greatest multiple of n in $[P_1^{\frac12}/(n-1)(n(n+1)/2)^{\frac12}] + n$, the difference (15) is $> \dfrac{P_1^{n-3/2}}{2(n(n+1)/2)^{\frac12}}$. When $A = n$, (15) is $\dfrac{n(n-1)}{2}P_1^{n-2}$.

The next inequality necessary is

$$(\text{II})\qquad \phi(P_1, An(n-1)) < \phi(P_1 + A(n-1)+1, 0) \qquad (0 \leqq A \leqq L_1 - 1),$$

where L_1 denotes $[n^{\frac12}P_1^{\frac12}/(n-1)(n(n+1)/2)^{\frac12}]$. The left side of (II) does not exceed

$$(16)\quad P_1^n + An(n-1)P_1^{n-1} + A^2n^2(n-1)^2P_1^{n-2} + (n+1)(n-1)^3A^3n^3P_1^{n-3} + \lambda,$$

where λ denotes a sum of terms in lower powers of P_1 with b_i $(i \geqq 4)$ having their maximum values given below (12). Replace the left side of (II) by (16) and subtract from the right side of (II). The difference exceeds

$$(17)\quad nP_1^{n-1} - P_1^{n-2}(n-1)^2(A^2n(n+1)/2 - An) - A^3n^3(n+1)(n-1)^3P_1^{n-3}.$$

This follows since the fourth term on the right side, which contains a higher power of P_1 than does λ, has been discarded along with the terms following it.

The difference (17) is minimum when A takes its maximum value $L_1 - 1$. When $A = L_1 - 1$, (17) is greater than $(n-1)n\left(\frac{n+1}{2}\right)^{\frac{1}{2}} P_1^{n-3/2}$.

Let y take the values 0, $n(n-1)$, $2n(n-1)$, $3n(n-1)$, $\cdots$. Let Let $\Phi(P_1, i) \equiv \phi(P_1, in(n-1))$. Let A in (I) be denoted by A_1. If y takes the value $jn(n-1)$ then A_1 must be a multiple jn of n. Thus (I) becomes

$$\text{(I}'\text{)}\quad \Phi(P_1, j) > \Phi(P_1 + (n-1), j-1) \qquad \left(j = 1, 2, \cdots, \frac{P_1}{n(n+1)(n-1)^2}\right).$$

Since in (II) the y-coördinates are multiples of n, we may write (II) as

$$\text{(II}'\text{)}\qquad \Phi(P_1, A) < \Phi(P_1 + A(n-1) + 1, 0) \qquad (1 \leqq A \leqq L_1 - 1).$$

By (I′)

$$\Phi(P_1 + A(n-1) + 1, 0) < \Phi(P_1 + (A-1)(n-1) + 1, 1)$$
$$< \cdots < \Phi(P_1 + (A-i)(n-1) + 1, i).$$

By (II′) and this

$$\Phi(P_1, A) < \Phi(P_1 + (A-i)(n-1) + 1, i) \qquad (i = 0, \cdots, A).$$

In the proof of (II) we found that the difference between the right and left sides exceeds $(n-1)n\left(\frac{n+1}{2}\right)^{\frac{1}{2}} P_1^{n-3/2}$. Hence the difference between the two sides of the last inequality is greater than $(n-1)in\left(\frac{n+1}{2}\right)^{\frac{1}{2}} P_1^{n-3/2}$ for all values of i. By (I′)

$$\Phi(P_1, A) > \Phi(P_1 + (n-1), A-1) > \cdots > \Phi(P_1 + k(n-1), A-k),$$

which becomes, when $k = A - i$,

$$\Phi(P_1, A) > \Phi(P_1 + (A-i)(n-1), i) \qquad (i = Q/n, Q/n + 1, \cdots, A).$$

We restrict i to begin at Q/n (where Q as we have stated above is the greatest multiple of n in $[P_1^{\frac{1}{2}}/(n-1)(n(n+1)/2)^{\frac{1}{2}}] + n$) in order that the difference between the two sides of the last inequality exceed

$$\frac{(n-1)}{2}\left(\frac{P_1^{n-3/2}}{(n(n+1)/2)^{\frac{1}{2}}(n-1)}\right)$$

for the values of i given. Thus

$$\text{(III)}\quad \begin{aligned} &\Phi(P_1 + (A-i)(n-1), i) < \Phi(P_1, A) \\ &\qquad < \Phi(P_1 + (A-i)(n-1) + 1, i) \qquad (i = Q/n, \cdots, A-1), \end{aligned}$$

and in which $A \leqq L_1 - 1$. Hence the values of $\phi(x, y)$ at the points

$$(P_1 + k, in(n-1)) \qquad (0 \leqq k \leqq P_1, Q/n \leqq i \leqq L_1 - 1)$$

will all be distinct. For, by (III) the value of $\phi(x, y)$ at any point in this rectangle lies between the values at two successive points in the same horizontal line determined by i, for every i. Also, in any horizontal line $\phi(x, y)$ is strictly increasing with x.

The number of points (and hence of distinct values of ϕ) in this rectangle is $(L_1 - Q/n)P_1 > (n^{\frac{1}{2}} - 1)P_1^{3/2}/(n - 1)\left(\dfrac{n(n + 1)}{2}\right)^{\frac{1}{2}}$. The difference between any two values of $\phi(x, y)$ in the rectangle exceeds $P_1^{n-3/2}/(2n(n + 1))^{\frac{1}{2}} = c_1 P_1^{n-3/2}$.

5. **The numbers ξ and u.** In the last part we determined for P_1 a set of points (x, y) at which $\phi(x, y)$ takes on distinct values. These points lie in the rectangle

$$P_1 \leqq x \leqq 2P_1, \qquad c_4 P_1^{\frac{1}{2}} \leqq y < c_0 P_1^{\frac{1}{2}},$$

where $c_4 P_1^{\frac{1}{2}} = [(1/n)[P_1^{\frac{1}{2}}/(n(n + 1)/2)^{\frac{1}{2}}(n - 1)] + 1](n - 1)$,

$$c_0 P_1^{\frac{1}{2}} = [n^{\frac{1}{2}} P_1^{\frac{1}{2}}/(n(n + 1)/2)^{\frac{1}{2}}(n - 1)]n(n - 1).$$

Let $P_i = [\gamma^{i-1} P_1^{(1-3\nu/2)^{i-1}}]$ $(i = 1, \cdots, k)$, where $\gamma = 1$ if $(c_1/kc_3)^\nu 2^{3\nu/2-1} > 1$, otherwise $\gamma = (c_1/kc_3)^\nu 2^{3\nu/2-2}$ $(c_3 = 2^n + 1)$, whence γ is independent of i and P_1.

For each P_i, as for P_1, a rectangle Δ_i can be determined containing a set of points (x, y) at which $\phi(x, y)$ takes on distinct values. As before, for each such set the difference between any two distinct values of $\phi(x, y)$ exceeds $c_1 P_i^{n-3/2}$. Also

$$\phi(x, y) \leqq \phi(2P_i, c_0 P_i^{\frac{1}{2}}) \leqq c_3 P_i^n.$$

Let (x_i, y_i) and (z_i, w_i) be in Δ_i and distinct. We shall prove

$$\phi(x_i, y_i) - \phi(z_i, w_i) > c_1 P_i^{n-3/2} > k\phi(2P_{i+1}, c_0 P_{i+1}^{\frac{1}{2}}) \geqq \sum_{i+1}^{k} \phi(x_j, y_j), \tag{18}$$

where (x_j, y_j) lie in Δ_j $(j > i)$.

The first inequality in (18) follows from the minimum difference. The rectangles are such that the maximum $\phi(x, y)$ with (x, y) in Δ_{i+1} is greater than the maximum of $\phi(x, y)$ with (x, y) chosen from any succeeding set Δ. Moreover the Δ_i $(i = 1, \cdots, k)$ are such that the second inequality holds. Note that the maximum value of $\phi(x_j, y_j)$ $(j \geqq i + 1)$ is attained at $(2P_j, c_0 P_j^{\frac{1}{2}})$ and there

$$\phi(x_l, y_l) < c_3 P_l^n < 2^n P_{i+1}^n < \phi(x_{i+1}, y_{i+1}) < c_3 P_{i+1}^n \qquad (l > i + 1).$$

Then

$$kc_3 P_{i+1}^n > \sum_{j=i+1}^{k} \phi(x_j, y_j),$$

and

$$k\phi(2P_{i+1}, c_0 P_{i+1}^{\frac{1}{2}}) < kc_3 P_{i+1}^n.$$

Hence the second inequality in (18) holds if

$$P_{i+1} < (c_1/kc_3)^{\nu} P_i^{1-3\nu/2}.$$

This will follow if

$$P_{i+1} \leqq \gamma^i P_1^{(1-3\nu/2)i},$$

and

$$P_i > \tfrac{1}{2}\gamma^{i-1} P_1^{(1-3\nu/2)i-1}$$

for P_1 large. Now

$$P_{i+1} < (c_1/kc_3)^{\nu} P_i^{1-3\nu/2}$$

if

$$\gamma^i P_1^{(1-3\nu/2)i} < (\tfrac{1}{2}\gamma^{i-1})^{1-3\nu/2}(c_1/kc_3)^{\nu} P_1^{(1-3\nu/2)i}.$$

That is, if

$$\gamma^{i-(i-1)(1-3\nu/2)} < (c_1/kc_3)^{\nu} 2^{3\nu/2-1} \qquad (i = 1, \cdots, k).$$

If the inequality is true when $i = 1$, it will hold for $i > 1$ provided $\gamma \leqq 1$. When $i = 1$, it is satisfied as follows. If $(c_1/kc_3)^{\nu} 2^{3\nu/2-1} \leqq 1$, $\gamma < (c_1/kc_3)^{\nu} 2^{3\nu/2-1}$, i.e. $\gamma = (c_1/kc_3)^{\nu} 2^{3\nu/2-2}$. If $(c_1/kc_3)^{\nu} 2^{3\nu/2-1} > 1$, $\gamma = 1$.

LEMMA.[4] *If we have an integer N such that*

$$N = \phi(x_1, y_1) + \phi(x_2, y_2) + \cdots + \phi(x_i, y_i) + \cdots + \phi(x_k, y_k),$$

with x_i, y_i in Δ_i then this representation of N is unique.

For, if also

$$N = \phi(x_1, y_1) + \cdots + \phi(x_{i-1}, y_{i-1}) + \phi(x_i', y_i') + \cdots + \phi(x_k', y_k'),$$

with x_i', y_i' in Δ: but distinct from x_i, y_i, then by (18)

$$\phi(x_i, y_i) > \phi(x_i', y_i') + \cdots + \phi(x_k', y_k').$$

Hence we reach a contradiction.

To form the numbers ξ, take all possible sums of k values of $\phi(x, y)$, one value chosen from each rectangle of sets (x, y). These are distinct by the last lemma. The number of ξ's will be the product of the number of values arising from each rectangle. Hence if X is the number of ξ's,

$$X > \chi(P_1^{3/2})^{\sum_{i=1}^{k}(1-3\nu/2)^i} = \chi P_1^{\frac{3}{2}\left(\frac{1-(1-3\nu/2)^k}{3\nu/2}\right)} = \chi P_1^{n-n(1-3\nu/2)^k},$$

where χ is a product of constants. Hence by (8)

$$X > \chi(n/2^{n+2})^{1-(1-3\nu/2)^k} P^{(n-1)(1-(1-3\nu/2)^k)},$$

which by (6) is

$$X > \chi(n/2^{n+2})^{1-(1-3\nu/2)^k} P^{n-1-\sigma}. \tag{19}$$

[4] Proved in a special case by H. Davenport and H. Heilbronn, *On Waring's theorem for fourth powers*, Proceedings of the London Mathematical Society, vol. 41 (1936), p. 143.

The integers u defined by

$$u = \xi + v^n \qquad (v = P, P + 1, \cdots, 2P - 1) \tag{20}$$

will be proved to be distinct. By (8)

$$nP^{n-1} \geqq P_1^n(2^{n+1}). \tag{21}$$

Also

$$(P + j)^n > P^n + nP^{n-1} \text{ for } j \geqq 1.$$

Now ξ maximum does not exceed $\phi(2P_1, c_0P_1^{\frac{1}{2}} + k2^{n+1}P_1^{n-3/2})$, and hence is less than $2^{n+1}P_1^n$. Hence since

$$(P + j)^n > (P^n + nP^{n-1}) \geqq P^n + 2^{n+1}P_1^n,$$

and

$$(P + 1)^n - P^n < (P + j)^n - (P + j - 1)^n \qquad j > 1,$$

the u's are all distinct. The maximum u is less than

$$(2P - 1)^n + 2^{n+1}P_1^n < (2P - 1)^n + nP^{n-1} < (2P)^n.$$

The minimum u is greater than P^n. Hence

$$P^n < u < (2P)^n. \tag{22}$$

Just as we defined the numbers ξ and u by use of P, P_1, we now define ξ_1 and u_1 by use of R, R_1. Under these replacements let χ become χ_1. Thus if X_1 is the number of ξ_1's, $X_1 > \chi_1(n/2^{n+2})^{1-(1-3\nu/2)^k}R^{n-1-\sigma}$. Hence

$$u_1 = \xi_1 + v_1^n \qquad (v_1 = R, \cdots, 2R - 1), \tag{23}$$

$$X_1 > \chi_1(n/2^{n+2})^{1-(1-3\nu/2)^k}R^{n-1-\sigma}, \tag{24}$$

$$R^n < u_1 < (2R)^n. \tag{25}$$

6. **The fundamental integral.** Take $N = N_0$ and consider

$$I_{N_0} = \sum_{y=1}^{Y} I_{yN_0} = \int_{-\tau^{-1}}^{1-\tau^{-1}} T_1 \cdots T_{s-3}TV^2S^2 \sum_y V_yS_ye^{-2\pi i\alpha N_0}\, d\alpha. \tag{88}$$

As in Dickson we obtain

$$I_{N_0} = H_1 + H_2, \qquad H_1 = \sum_y H_{y1} > c_9YX^2X_1RP^{s-n}, \tag{89}$$

$$| H_2 | < DP^{s-2}PX(X_1RYP^{z+n-\frac{1}{2}})^{\frac{1}{2}}, \qquad D = 3(8mn)^{\frac{1}{2}}, \tag{96}$$

with his k replaced by $k + 1$. The constants arise in the same way as in his paper with values appropriate to this paper. The formulas are numbered as in his paper. By (7), (9), (19), (24),

$$R > \sqrt{\tfrac{1}{2}}\, P^{1-f}, \qquad Y > \sqrt{\tfrac{1}{2}}\, C^{\nu}P^{(1-\nu)f}, \tag{97}$$

$$X \geqq (n/2^{n+2})^{1-(1-3\nu/2)^k}\chi P^{n-1-\sigma}$$

while X_1 exceeds the last products with P replaced by R and χ replaced by χ_1. Note that

$$(98) \qquad \sigma = (n-1)(1-3\nu/2)^k, \qquad \mu = 1-(1-3\nu/2)^k.$$

Multiply and divide (96) by $X(X_1RY)^{\frac{1}{2}}$ and apply (97). Thus

$$(99) \qquad |H_2| < C_1P^JX^2X_1RYP^{s-n}, \qquad C_1 = \frac{2^{(n+1-\sigma)/4}D}{\chi(n/2^{n+2})^{\mu}\{C^{\nu}\chi_1(n/2^{n+2})^{\mu}\}^{\frac{1}{2}}},$$

$$(100) \qquad 2J = \sigma(3-f) - g, \qquad g = \tfrac{1}{2} - z - nf + (1-\nu)f,$$

The Waring theorem is true for every integer $\geqq N_0$ if the integral $I_{N_0} > 0$. By the above and (89), this is true if

$$(101) \qquad P^{-J} > C_1/c_9\,.$$

For P large, this holds if $J < 0$. This is true by (98_2) if

$$(102) \qquad r(1-3\nu/2)^k < 1, \qquad r = (n-1)(3-f)/g.$$

But r increases with f since

$$\frac{dr}{df} = \left(\frac{n-1}{g^2}\right)(3n - 7/2 + z + 3\nu).$$

Since (102_1) is equivalent to

$$(103) \qquad k > \frac{\log r}{-\log(1-3\nu/2)},$$

k will be minimum if r and hence f is. But $f \geqq \nu/2$. Hence we employ the minimum f, viz.,

$$(104) \qquad f = \nu/2.$$

The resulting value of r in (102) is

$$(105) \qquad r = n^2(6n - 7 + \nu)/(n-d), \qquad d = 1 + 2n^2z.$$

For a small z, say $\nu^3/24$, let k_0 be the least integer k satisfying (103). Then all sufficiently large integers are sums of $4n - 2 + 3k_0$ values of a homogeneous polynomial of the type previously described.

THE UNIVERSITY OF CHICAGO.

ANNALS OF MATHEMATICS
Vol. 39, No. 1, January, 1938

ON A CERTAIN CLASS OF NON-LINEAR EXPANSIONS OF AN ARBITRARY ANALYTIC FUNCTION

BY GERTRUDE S. KETCHUM AND P. W. KETCHUM

(Received April 3, 1937)

1. **Introduction.** In these Annals, vol. 33 (1932) pp. 139–42, J. M. Feld proved the following theorem:

Let $b_1, b_2, \cdots$ be a given sequence of complex numbers such that

$$\text{(a)} \qquad |b_{n+1}| \geqq |b_n| \geqq 1,$$

$$\text{(b)} \qquad |b_n|^{1/n} \text{ is bounded for all } n.$$

Then every function $f(x) = \Sigma c_n x^n$, analytic at the origin, can be expanded uniquely in some neighborhood of the origin in an absolutely and uniformly convergent series of the form

$$f(x) = f(0) + \sum_1^{\infty} b_n \frac{a_n x^n}{1 - a_n x^n}. \tag{1}$$

In the present note we generalize this theorem to include expansions of the form

$$f(x) = f(0) + \sum_1^{\infty} b_n F_n(a_n x^n). \tag{2}$$

The $F_n(w)$ are given functions, each of which has a simple zero at the origin and is, after a certain normalization, analytic and bounded in some fixed circular region as follows:

$$F_n(w) = \sum_{s=1}^{\infty} \alpha_{n,s} w^s, \qquad \alpha_{n,1} = 1, \qquad \text{for } |w| \leqq \rho, \tag{3}$$

$$|F_n(w)| \leqq M_n \text{ for } |w| \leqq \rho, \quad \text{where} \quad M_n = 0(n^a) \text{ for some } a. \tag{4}$$

Also, condition (a) may be replaced by the more general condition that there exist constants C, a, and σ so large that

$$\text{(a}') \qquad 0 < |b_{n_1}| \leqq C n^a \sigma^{n/n_1} |b_n|$$

for all positive integers n and all integers n_1 which are divisors of n less than n. Finally, we obtain an estimate for the region of validity of (2) which for the particular case (1) is superior to Feld's estimate. This is accomplished by replacing the bound in (b) by

$$\text{(b}') \qquad \limsup_{n \to \infty} |b_n|^{1/n} = K < \infty,$$

by using a better estimate than Feld's for the number of divisors of an integer, and by making use of a table of values of the a's for small values of n.

Problems similar to the above in which the a's rather than the b's are given have received considerable attention in the literature.[1] In any case, regardless of whether the a's or the b's are given, the expansion problem can be related to an associated problem on the solution of an infinite system of recurrence relations on the unknown b's or a's respectively. This system is obtained by formally expanding $f(x)$ and $F_n(w)$ in powers of x in (2) and equating coefficients of like powers of x. The distinguishing feature of the present problem is that these recurrence relations are *non-linear*. Thus one cannot obtain the expansion (of the form (2)) of $f(x) + g(x)$ by simply adding the expansions of $f(x)$ and $g(x)$, nor can one obtain the expansion of this form for $2f(x)$ by doubling the expansion of $f(x)$.

As a corollary to our general theorem, we prove that if $b_1, b_2, \cdots$ are given satisfying conditions (a′) and (b), and if $g_n(w)$ are given functions such that the ratios $wg_n'(w)/(1 + g_n(w))$ satisfy the conditions imposed on $F_n(w)$ in (3) and (4), then any function analytic and non-vanishing at the origin can be expanded in the infinite product

$$f(x) = f(0) \prod_1^{\infty} (1 + g_n(a_n x^n))^{b_n/n}, \qquad (g_n(0) = 0). \tag{5}$$

Particular cases have been studied previously by Feld and by Ritt.[2]

2. **Certain Number-Theoretic Lemmas.** We first introduce certain numerical functions on which the subsequent work depends. Let ν be a given positive integer or zero. Let $\phi(p, s)$ be a function of the positive integers p and s where s is a divisor of p less than p. Let

$$\psi_\nu\{\phi\} = \underset{n, n_1, \cdots, n_{k+1}}{\text{l.u.b.}} [| \phi(n, n_1) |^{1/n} | \phi(n_1, n_2) |^{1/n_1} \cdots | \phi(n_k, n_{k+1}) |^{1/n_k}], \tag{6}$$

where the least upper bound is taken over all permissible values of

$$n, n_1, \cdots, n_{k+1},$$

with $n_k \geqq 2^\nu$. It is understood that there may be only one factor inside the brackets, under the convention that $n_0 = n$. We are only interested in those functions $\phi(n, n_1)$ for which $\psi_\nu\{\phi\}$ is finite. A sufficient condition for this is contained in the following lemma.

Lemma I. *If $\phi(n, n_1)$ is such that for sufficiently large positive constants C, a, and σ,*

$$| \phi(n, n_1) | \leqq Cn^a \sigma^{n/n_1}, \tag{7}$$

then $\psi_\nu\{\phi\}$ is finite. On the other hand, if in this condition $C, \sigma \geqq 1$ and the factor n^a is replaced by e^{n^ϵ} or σ^{n/n_1} by $n^{\epsilon n/n_1}$, $\epsilon > 0$, then, no matter how small ϵ

[1] See Gertrude S. Ketchum, Trans. Am. Math. Soc., **40** (1936), pp. 208–24.

[2] J. F. Ritt, Math. Zeit., **32** (1930), pp. 1–3.

may be, the condition so modified will be satisfied by some $\phi(n, n_1)$ *for which* $\psi_\nu\{\phi\}$ *is infinite.*

We have the following three relationships:

$$\psi_\nu\{\phi_1\phi_2\} \leqq \psi_\nu\{\phi_1\}\psi_\nu\{\phi_2\}, \tag{8}$$

$$\psi_\nu\{\phi^a\} = [\psi_\nu\{\phi\}]^a \text{ if } a \geqq 0, \tag{9}$$

$$\psi_\nu\{\Phi\} \leqq \psi_\nu\{\phi\} \text{ whenever } |\Phi| \leqq |\phi|. \tag{10}$$

The first part of the lemma follows from these three relations and the following inequalities:

$$\psi_\nu\{C\} \leqq [\operatorname{Max}(1, C)]^{1/2^{\nu-1}}, \tag{11}$$

$$\psi_\nu\{n\} \leqq 2^{2-(1/2)-(2/2^2)-\cdots-((\nu-1)/2^{\nu-1})}, \tag{12}$$

$$\psi_\nu\{\sigma^{n/n_1}\} \leqq \psi_0\{\sigma\}. \tag{13}$$

Of these, (11) and (13) are elementary and (12) may be obtained by a slight modification of Feld's proof of a corresponding result.

The second part of the lemma follows from consideration of the functions $\psi_\nu\{e^{\epsilon n}\}$ and $\psi_\nu\{n^{\epsilon n/n_1}\}$.

LEMMA II. *If* $F_1(w)$, $F_2(w)$, $\cdots$ *is a sequence of functions satisfying conditions* (3) *and* (4), *then* $\psi_\nu\{\alpha_{n_1,n/n_1}\}$ *is finite.*

Proof: By Cauchy's inequality,

$$|\alpha_{n,n/n_1}| \leqq M_n/\rho^{n/n_1} \leqq An^a(1/\rho)^{n/n_1}.$$

Hence by (8), (9), and (10),

$$\psi_\nu\{\alpha_{n,n/n_1}\} \leqq \psi_\nu\{A\}[\psi_\nu\{n\}]^a\psi_\nu\{(1/\rho)^{n/n_1}\}. \tag{14}$$

3. **Proof of Main Theorem.** Let $F_1(w)$, $F_2(w)$, $\cdots$ be given functions satisfying (3) which are such that $\psi_\nu\{\alpha_{n,\,n/n_1}\}$ is finite. (If the given functions $F_n(w)$ do not satisfy the condition $\alpha_{n,1} = 1$, we may replace them by a new set of functions $\bar{F}_n(w) = F_n(w)/\alpha_{n,1}$ and the given constants b_n by a new set $\bar{b}_n = b_n\alpha_{n,1}$. Having found an expansion of $f(x)$ in terms of these new functions and constants in a certain region S, there will be determined an expansion of $f(x)$ in the original set valid in the same region S.) Let b_1, b_2, $\cdots$ be given constants satisfying (b) and

$$\text{(a}''\text{)} \qquad b_n \neq 0, \qquad \psi_\nu\{b_{n_1}/b_n\} < \infty, \qquad b_1 = 1.$$

(The last condition $b_1 = 1$ is introduced for convenience of notation. If b_1 is not equal to unity we may replace $f(x)$ by the new function $\bar{f}(x) = f(x)/b_1$ and the given constants by the new set $\bar{b}_n = b_n/b_1$. Having found an expansion in S for $\bar{f}(x)$ with the new set of b's, on multiplying the terms of the expansion by b_1 there will be obtained an expansion of $f(x)$ with the original b's,

valid in the same region S.) Consider the formal equality (2) where $f(x)$ is analytic at $x = 0$, so that for $n \geqq 2^\nu$, $|c_n|^{1/n}$ is bounded by some constant r_ν; and where the a's are defined by the recurrence relations

$$c_1 = a_1$$

$$(15) \qquad c_n = \sum_{k=0}^{t} b_{d_k} \alpha_{d_k, n/d_k} a_{d_k}^{n/d_k} + b_n a_n, \qquad (n = 2, 3, \cdots).$$

Here $d_0, d_1, \cdots, d_t$ are all the divisors of n less than n, arranged in order of magnitude beginning with $d_0 = 1$.

We seek a bound for $|a_n|^{1/n}$. It is desired to make this bound as small as is practicable in order to insure a good estimate of the region of validity of (2). Since the bound which we obtain is affected to a considerable extent by the values of $|a_n|$ for small values of n, and since it is possible to compute a table of such values for small n, our results are stated in such a way as to enable one to utilize such a table. Thus, let L_ν denote a bound for $|a_n|^{1/n}$ for $n = 1, 2, \cdots, 2^\nu - 1$ where ν may be 1 or any other positive integer as desired. We will now obtain a bound for $|a_n|^{1/n}$ for $n \geqq 2^\nu$. We have

$$|a_1| = |c_1|$$

$$(16) \qquad |a_n| \leqq \frac{|c_n|}{|b_n|} + \sum_{k=0}^{t} \frac{|b_{d_k}|}{|b_n|} |\alpha_{d_k, n/d_k} a_{d_k}^{n/d_k}|, \qquad (n = 2, 3, \cdots).$$

In the term under the summation whose absolute value is largest, let us write $d_k = n_1$. There can be at most $d(n) - 1$ terms under the summation, where $d(n)$ is the number of divisors of n. Hence

$$(17) \qquad |a_n| \leqq \frac{|c_n|}{|b_n|} + (d(n) - 1) \frac{|b_{n_1}|}{|b_n|} |\alpha_{n_1, n/n_1} a_{n_1}^{n/n_1}|.$$

If

$$(18) \qquad \left|\frac{c_n}{b_n}\right| \geqq |\alpha_{n_1, n/n_1} a_{n_1}^{n/n_1}| \left|\frac{b_{n_1}}{b_n}\right|,$$

then

$$(19) \qquad |a_n|^{1/n} \leqq [d(n)]^{1/n} \left|\frac{1}{b_n}\right|^{1/n} |c_n|^{1/n}.$$

If, on the other hand, the sign of inequality in (18) is reversed, then

$$(20) \qquad |a_n|^{1/n} \leqq [d(n)]^{1/n} |\alpha_{n, n/n_1}|^{1/n} \left|\frac{b_{n_1}}{b_n}\right|^{1/n} |a_{n_1}|^{1/n_1}.$$

If (20) holds and $n_1 \geqq 2^\nu$, we treat $|a_{n_1}|^{1/n_1}$ just as we have treated $|a_n|^{1/n}$. Proceeding in this way until the process stops of itself or until after $k + 1$

steps $n_{k+1} < 2^\nu$ while $n_k \geqq 2^\nu$, we find that $|a_n|^{1/n}$ is bounded as in (19) or by one of the two following expressions

$$(21)\quad \psi_\nu\{d(n)\} \, |\alpha_{n_1,n/n_1}|^{1/n} \cdots |\alpha_{n_k,n_{k-1}/n_k}|^{1/n_{k-1}} \left|\frac{b_{n_1}}{b_n}\right|^{1/n} \cdots \left|\frac{b_{n_k}}{b_{n_{k-1}}}\right|^{1/n_{k-1}} \left|\frac{1}{b_{n_k}}\right|^{1/n_k} |c_{n_k}|^{1/n_k},$$

$$(22)\quad \psi_\nu\{d(n)\} \, |\alpha_{n_1,n/n_1}|^{1/n} \cdots |\alpha_{n_{k+1},n_k/n_{k+1}}|^{1/n_k} \left|\frac{b_{n_1}}{b_n}\right|^{1/n} \cdots \left|\frac{b_{n_{k+1}}}{b_{n_k}}\right|^{1/n_k} |a_{n_{k+1}}|^{1/n_{k+1}}.$$

These three expressions in (19), (21), and (22) may then be replaced by their least upper bounds. After so doing, the number 1 in the numerator of $|1/b_{n_k}|$ may be replaced by $b_{n_{k+1}}$. Hence,

$$(23)\quad \begin{gathered} |a_n|^{1/n} \leqq N, \\ N = \psi_\nu\{d(n)\}\, \psi_\nu\left\{\frac{b_{n_1}}{b_n}\right\} \operatorname{Max}(1, \psi_\nu\{\alpha_{n_1,n/n_1}\}) \operatorname{Max}(r_\nu, L_\nu). \end{gathered}$$

We proceed now to the convergence of the series (2). By Schwarz' lemma, if the functions $F_n(w)$ satisfy (4), then

$$|F_n(a_n x^n)| \leqq An^a\, |a_n|\, |x|^n/\rho \quad \text{for} \quad |a_n x^n| \leqq \rho,$$

and hence certainly for $|x| \leqq \rho^{1/n}/N$. For $|x| < 1/N$ and $n \geqq m$, where m is sufficiently large, the series in (2) is dominated by

$$\sum_{n=m}^{\infty} An^a N^n (K + \epsilon)^n \, |x|^n/\rho.$$

Thus (2) converges absolutely and uniformly for

$$(24)\quad |x| \leqq R < \operatorname{Min}\left(\frac{1}{NK}, \frac{1}{N}\right).$$

Moreover, by virtue of the recurrence relations (15) it follows from the Weierstrass double series theorem that the series converges in this region to $f(x)$. We have, therefore, proved

THEOREM I. *If $F_1(w)$, $F_2(w)$, $\cdots$ are given functions satisfying* (3) *and* (4) *and if b_1, b_2, $\cdots$ are given constants satisfying* (b) *and* (a″), *then any function $f(x)$ which is analytic at the origin has a unique expansion of the form* (2) *in an absolutely and uniformly convergent series in the region* (24). *In particular, a sufficient condition for* (a″) *is* (a′) *with $b_1 = 1$, in which case an estimate for N may be computed by means of formulas* (8)–(14).

4. **Application to Feld's Theorem.** Feld's theorem, stated in the Introduction, is obtained as an immediate consequence of our Theorem I on taking

$F_n(w) = w/(1 - w)$. If we take $\nu = 7$ and compute successive bounds for $|a_n|$ from the relations

$$|a_1| = |c_1/b_1|, \qquad |a_n| \leqq |c_n/b_1| + \sum_{k=0}^{t} |a_{d_k}^{n/d_k}|,$$

we get $L_7 \leqq |a_{96}|^{1/96} r^{(7)} < 1.7545 r^{(7)}$ where $r^{(\nu)}$ is the maximum of $|c_n/b_1|^{1/n}$ for $n = 1, 2, \cdots, 2^\nu - 1$. Since $d(n) \leqq 2n^{\frac{1}{2}}$, we have from equations (8)–(12), $\psi_7\{d(n)\} \leqq 2^{5/64}$. Hence the expansion of $f(x)$ is valid in a region about the origin not smaller than

$$|x| \leqq R < [2^{5/64} K \operatorname{Max}(r_7, 1.7545 r^{(7)})]^{-1} < (1.853 rK)^{-1},$$

where we have written $r = r_0$, and, due to the necessary normalization, r_ν is now a bound for $|c_n/b_1|^{1/n}$ for all $n \geqq 2^\nu$.

The region of validity obtained by Feld is similar to the last expression, but the numbers r and K are replaced by other quantities greater than or equal to them and the factor 1.853 is replaced by 8.

Let λ denote the greatest lower bound of all numbers σ such that the expansion of every function $f(x)$ is valid for $|x| < (\sigma r K)^{-1}$.

5. A Lower Bound for the Constant λ: Consider the expansion of the particular function $1/(x - 1)$ in Feld's series with $b_n \equiv 1$. Then $K = r = 1$. Hence $\lambda \geqq 1/|x_0|$ if the expansion diverges at x_0. But the series diverges for $|x| > [\limsup |a_n|^{1/n}]^{-1}$. For our particular case, $a_n = 0$ if n is odd and larger than 1. On the other hand, $|a_{2^k}|^{1/2^k}$ is always increasing with k, and approaches the limit $1.7355 \cdots$. Thus, $\limsup |a_n|^{1/n} \geqq 1.7355 \cdots$, and we have

THEOREM II. *The expansion asserted by Feld, in his theorem stated at the beginning of this paper, is valid for* $|x| < (\lambda\, r\, K)^{-1}$, *where* r *is a bound for* $|c_n/b_1|^{1/n}$ *for all* n, K *is defined as in* (b′), *and* λ *is a constant whose value is such that*

$$1.7355 \cdots \leqq \lambda < 1.853.$$

6. Expansions in Infinite Products. We have

THEOREM III. *Let* $g_1(w), g_2(w), \cdots$ *be given functions which are analytic for* $|w| \leqq \rho$ *and such that*

$$g_n(0) = 0, \qquad g_n'(0) = \alpha_{n,1} \neq 0, \qquad \left| \frac{g_n'(w)}{\alpha_{n,1}(1 + g_n(w))} \right| \leqq M_n$$

for $|w| \leqq \rho$, *where* $M_n = 0(n^a)$. *Let* $b_1, b_2, \cdots$ *be given constants satisfying conditions* (b) *and the first two parts of* (a″). *Then any function* $f(x)$ *which is analytic in the neighborhood of the origin and non-vanishing at the origin may be expanded in an absolutely and uniformly convergent product of the form* (5) *valid in some neighborhood of the origin.*

This theorem is obtained as a corollary to Theorem I on taking

$$F_n(w) = \frac{wg'_n(w)}{\alpha_{n,1}(1 + g_n(w))}.$$

The particular case $g_n(w) = -w$ satisfies all the required conditions so that we get

$$f(x) = f(0) \prod_1^\infty (1 - a_n x^n)^{b_n/n},$$

which is a product expansion of Feld. The particular case $b_n = n$ was studied by Ritt.

Another simple example is given by $g_n(w) = \sin w$ in which case we have the expansion

$$f(x) = f(0) \prod_1^\infty (1 + \sin a_n x^n)^{b_n/n}.$$

INSTITUTE FOR ADVANCED STUDY.

Annals of Mathematics
Vol. 39, No. 1, January, 1938

THE GRAVITATIONAL EQUATIONS AND THE PROBLEM OF MOTION

By A. Einstein, L. Infeld, and B. Hoffmann

(Received June 16, 1937)

Introduction. In this paper we investigate the fundamentally simple question of the extent to which the relativistic equations of gravitation determine the motion of ponderable bodies.

Previous attacks on this problem[1] have been based upon gravitational equations in which some specific energy-momentum tensor for matter has been assumed. Such energy-momentum tensors, however, must be regarded as purely temporary and more or less phenomenological devices for representing the structure of matter, and their entry into the equations makes it impossible to determine how far the results obtained are independent of the particular assumption made concerning the constitution of matter.

Actually, the only equations of gravitation which follow without ambiguity from the fundamental assumptions of the general theory of relativity are the equations for empty space, and it is important to know whether they *alone* are capable of determining the motion of bodies. The answer to this question is not at all obvious. It is possible to find examples in classical physics leading to either answer, yes or no. For instance, in the ordinary Maxwell equations for empty space, in which electrical particles are regarded as point singularities of the field, the motion of these singularities is not determined by the linear field equations. On the other hand, the well-known theory of Helmholtz on the motion of vortices in a non-viscous fluid gives an instance where the motion of line singularities is actually determined by partial differential equations alone, which are there non-linear.

We shall show in this paper that the gravitational equations for empty space are in fact sufficient to determine the motion of matter represented as point singularities of the field. The gravitational equations are non-linear, and, because of the necessary freedom of choice of the coördinate system, are such that four differential relations exist between them so that they form an overdetermined system of equations. The overdetermination is responsible for the existence of equations of motion, and the non linear character for the existence of terms expressing the interaction of moving bodies.

Two essential steps lead to the determination of the motion.

---

[1] Droste, Ac. van Wet. Amsterdam *19*, 447 (1916). De Sitter, Monthly Notices of the R. A. S. *67*, 155 (1916). Mathisson, Zeits. f. Physik, *67*, 270, 826 (1931), *69*, 389 (1931). Levi-Civita, Am. Jour. of Math., lix, 3, 225 (1937).

(1) By means of a new method of approximation, specially suited to the treatment of quasi-stationary fields, the gravitational field due to moving particles is determined.

(2) It is shown that for two-dimensional spatial surfaces containing singularities certain surface integral conditions are valid which determine the motion.

In the second part of this paper we actually calculate the first two non-trivial stages of the approximation. In the first of these the equations of motion take the Newtonian form. In the second the equations of motion, which we calculate only for the case of two massive particles, take a more complicated form but do not involve third or higher derivatives with respect to the time.

The method is, in principle, applicable for any order of approximation, the problem reducing to specific integrations at each stage, but we have not proved that higher time derivatives than the second will not ultimately occur in the equations of motion.

In the determination of the field and the equations of motion non-Galilean values at infinity and singularities of the type of dipoles, quadrupoles, and higher poles, must be excluded from the field in order that the solution shall be unique.

It is of significance that our equations of motion do not restrict the motion of the singularities more strongly than the Newtonian equations, but this may be due to our simplifying assumption that matter is represented by singularities, and it is possible that it would not be the case if we could represent matter in terms of a field theory from which singularities were excluded. The representation of matter by means of singularities does not enable the field equations to fix the sign of mass so that, so far as the present theory is concerned, it is only by convention that the interaction between two bodies is always an attraction and not a repulsion. A possible clue as to why the mass must be positive can be expected only from a theory which gives a representation of matter free from singularities.[2]

Our method can be applied to the case when the Maxwell energy-momentum tensor is included in the field equations and, as is shown in part II, it leads to a derivation of the Lorentz force.

In the Maxwell-Lorentz electrodynamics, as also in the earlier approximation method for the solution of the gravitational equations, the problem of determining the field due to moving bodies is solved through the integration of wave equations by retarded potentials. The sign of the flow of time there plays a decisive rôle since, in a certain sense, the field is expanded in terms of only those waves which proceed towards infinity. In our theory, however, the equations to be solved at each stage of the approximation are not wave equations but merely spatial potential equations. Since such equations as those of the gravitational and of the electromagnetic field are actually invariant under a

[2] Einstein and Rosen, Phys. Rev. vol. 48, 73 (1935).

reversal of the sign of time, it would seem that the method presented here, is the natural one for their solution. Our method, in which the time direction is not distinguished, corresponds to the introduction of standing waves in the wave equation and cannot lead to the conclusion that in the circular motion of two point masses energy is radiated to infinity in the form of waves.

I. General Theory

1. **Field Equations and Coördinate Conditions.** Since it is an essential part of the work to make a separation between space and time we shall, throughout this paper, use the convention that Latin indices take on only the spatial values 1, 2, 3 while Greek indices refer to both space and time, running over the values 0, 1, 2, 3.

As explained in the introduction, we discuss only the gravitational equations for empty space, treating the sources of the field as singularities. If we denote the ordinary derivative of a quantity by means of a line followed by the appropriate suffix, as

$$(1,1) \qquad \frac{\partial g_{\mu\nu}}{\partial x^\sigma} \rightarrow g_{\mu\nu|\sigma}\,; \qquad \frac{\partial^2 g_{\mu\nu}}{\partial x^\sigma\,\partial x^\rho} \rightarrow g_{\mu\nu|\sigma\rho}\,,$$

we may write the field equations in the form

$$(1,2) \qquad R_{\mu\nu} = -\begin{Bmatrix}\lambda\\ \mu\nu\end{Bmatrix}_{|\lambda} + \begin{Bmatrix}\lambda\\ \mu\lambda\end{Bmatrix}_{|\nu} + \begin{Bmatrix}\lambda\\ \mu\sigma\end{Bmatrix}\begin{Bmatrix}\sigma\\ \lambda\nu\end{Bmatrix} - \begin{Bmatrix}\lambda\\ \mu\nu\end{Bmatrix}\begin{Bmatrix}\sigma\\ \lambda\sigma\end{Bmatrix} = 0.$$

Let the symbols $\eta_{\mu\nu}$, $\eta^{\mu\nu}$ be defined by

$$(1,3) \qquad \eta_{\mu\nu} = \eta^{\mu\nu} = \begin{Bmatrix} +1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{Bmatrix},$$

so that they represent the metric of empty space-time. Then if we introduce the quantities $h_{\mu\nu}$, $h^{\mu\nu}$ by the relations

$$(1,4) \qquad g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}\,, \qquad g^{\mu\nu} = \eta^{\mu\nu} + h^{\mu\nu},$$

the $h_{\mu\nu}$ and $h^{\mu\nu}$ will represent the deviation of space-time from the flat case. The $h^{\mu\nu}$ can be calculated as functions of the $h_{\mu\nu}$ by means of the relations

$$(1,5) \qquad g_{\mu\nu}g^{\nu\sigma} = \delta^\sigma_\mu\,.$$

In general the $h_{\mu\nu}$ will be small relative to unity, but we make no assumptions here concerning their order of magnitude.

By means of (1, 4), (1, 5) we can express the components of $R_{\mu\nu}$ as functions of the $h_{\mu\nu}$, and for reasons which will become clear when we come to the method of approximation used in the present work we separate the various terms so

obtained into two groups in the following manner. First we separate the terms linear in the h's from those which are quadratic and of higher order. At this stage of the separation the field equations are of the form

$$R_{00} = \tfrac{1}{2}\{-h_{00|ss} + 2h_{0s|0s} - h_{ss|00}\} + L'_{00} = 0, \tag{1, 6}$$

$$R_{0n} = \tfrac{1}{2}\{-h_{0n|ss} + h_{0s|ns} + h_{ns|0s} - h_{ss|n0}\} + L'_{0n} = 0, \tag{1, 7}$$

$$\begin{aligned} R_{mn} = \tfrac{1}{2}\{-h_{mn|ss} + h_{ms|ns} + h_{ns|ms} - h_{ss|mn} + h_{mn|00} - h_{m0|n0} \\ - h_{n0|m0} + h_{00|mn}\} + L'_{mn} = 0, \end{aligned} \tag{1, 8}$$

where the $L'_{\mu\nu}$ represent the non-linear terms. We now take

(1, 9) from R_{00} the terms $h_{0s|0s} - \tfrac{1}{2}h_{ss|00}$,

(1, 10) from R_{0n} no terms

(1, 11) and from R_{mn} the terms $-\tfrac{1}{2}h_{0m|0n} - \tfrac{1}{2}h_{0n|0m} + \tfrac{1}{2}h_{mn|00}$,

and add them to the non-linear group. Introducing the symbol $L_{\mu\nu}$ to denote the non-linear group $L'_{\mu\nu}$ together with these added linear members, we may write the field equations in the separated from

$$R_{00} = -\tfrac{1}{2}h_{00|ss} + L_{00} = 0, \tag{1, 12}$$

$$\begin{aligned} R_{0n} = -\tfrac{1}{2}h_{0n|ss} + \tfrac{1}{2}(h_{ns} - \tfrac{1}{2}\delta_{ns}h_{ll} + \tfrac{1}{2}\delta_{ns}h_{00})_{|0s} \\ - \tfrac{1}{4}(h_{00} + h_{ss})_{|0n} + \tfrac{1}{2}h_{0s|ns} + L_{0n} = 0, \end{aligned} \tag{1, 13}$$

$$\begin{aligned} R_{mn} = -\tfrac{1}{2}h_{mn|ss} + \tfrac{1}{2}(h_{ms} - \tfrac{1}{2}\delta_{ms}h_{ll} + \tfrac{1}{2}\delta_{ms}h_{00})_{|ns} \\ + \tfrac{1}{2}(h_{ns} - \tfrac{1}{2}\delta_{ns}h_{ll} + \tfrac{1}{2}\delta_{ns}h_{00})_{|ms} + L_{mn} = 0, \end{aligned} \tag{1, 14}$$

where the L's are given explicitly by the formulas

$$\begin{aligned} L_{00} = h_{0s|0s} - \tfrac{1}{2}h_{ss|00} - (h^{\lambda\sigma}[00, \sigma])_{|\lambda} + (h^{\lambda\sigma}[0\lambda, \sigma])_{|0} \\ + \begin{Bmatrix}\lambda \\ 0\sigma\end{Bmatrix}\begin{Bmatrix}\sigma \\ \lambda 0\end{Bmatrix} - \begin{Bmatrix}\lambda \\ 00\end{Bmatrix}\begin{Bmatrix}\sigma \\ \lambda\sigma\end{Bmatrix}, \end{aligned} \tag{1, 15}$$

$$L_{0n} = -(h^{\lambda\sigma}[0n, \sigma])_{|\lambda} + (h^{\lambda\sigma}[n\lambda, \sigma])_{|0} + \begin{Bmatrix}\lambda \\ n\sigma\end{Bmatrix}\begin{Bmatrix}\sigma \\ \lambda 0\end{Bmatrix} - \begin{Bmatrix}\lambda \\ n0\end{Bmatrix}\begin{Bmatrix}\sigma \\ \lambda\sigma\end{Bmatrix}, \tag{1, 16}$$

$$\begin{aligned} L_{mn} = -\tfrac{1}{2}h_{0m|0n} - \tfrac{1}{2}h_{0n|0m} + \tfrac{1}{2}h_{mn|00} - (h^{\lambda\sigma}[mn, \sigma])_{|\lambda} \\ + (h^{\lambda\sigma}[m\lambda, \sigma])_{|n} + \begin{Bmatrix}\lambda \\ m\sigma\end{Bmatrix}\begin{Bmatrix}\sigma \\ \lambda n\end{Bmatrix} - \begin{Bmatrix}\lambda \\ mn\end{Bmatrix}\begin{Bmatrix}\sigma \\ \lambda\sigma\end{Bmatrix}. \end{aligned} \tag{1, 17}$$

If we introduce the quantities $\gamma_{\mu\nu}$ defined by

$$\gamma_{\mu\nu} = h_{\mu\nu} - \tfrac{1}{2}\,\eta_{\mu\nu}\eta^{\sigma\rho}h_{\sigma\rho}, \tag{1, 18}$$

or, in expanded form,

$$\gamma_{00} = \tfrac{1}{2}h_{00} + \tfrac{1}{2}h_{ll}\,, \tag{1, 19}$$

$$\gamma_{0n} = h_{0n}\,, \tag{1, 20}$$

$$\gamma_{mn} = h_{mn} - \tfrac{1}{2}\delta_{mn}h_{ll} + \tfrac{1}{2}\delta_{mn}h_{00}\,, \tag{1, 21}$$

we may write the field equations (1, 12), (1, 13), (1, 14) in the form

$$R_{00} = -\tfrac{1}{2}h_{00|ss} + L_{00} = 0, \tag{1, 22}$$

$$R_{0n} = -\tfrac{1}{2}h_{0n|ss} + \tfrac{1}{2}\gamma_{ns|0s} + \tfrac{1}{2}(\gamma_{0s|s} - \gamma_{00|0})_{|n} + L_{0n} = 0, \tag{1, 23}$$

$$R_{mn} = -\tfrac{1}{2}h_{mn|ss} + \tfrac{1}{2}\gamma_{ms|sn} + \tfrac{1}{2}\gamma_{ns|sm} + L_{mn} = 0. \tag{1, 24}$$

Since there are four identities between these field equations, we may impose four coördinate conditions, in the form of four non-tensorial equations involving the gravitational potentials, so as to limit the arbitrariness of the solutions by limiting the freedom of choice of the coördinate system. It turns out to be simplest to use coördinate conditions which involve only quantities which enter the explicitly written parts of the field equations (1, 23), (1, 24). These equations, in fact, suggest that we take as our coördinate conditions[3]

$$\gamma_{0s|s} - \gamma_{00|0} = 0, \tag{1, 25}$$

$$\gamma_{ms|s} = 0. \tag{1, 26}$$

With these coördinate conditions the field equations become merely

$$h_{00|ss} = 2L_{00}\,, \tag{1, 27}$$

$$h_{0n|ss} = 2L_{0n}\,, \tag{1, 28}$$

$$h_{mn|ss} = 2L_{mn}\,. \tag{1, 29}$$

For the further argument it is necessary that we write these equations in such a way that the Laplacians of the γ's enter instead of the Laplacians of the h's. We therefore replace the above equations by the equivalent equations

$$\gamma_{00,ss} = 2\Lambda_{00}\,, \tag{1, 30}$$

$$\gamma_{0n,ss} = 2\Lambda_{0n}\,, \tag{1, 31}$$

$$\gamma_{mn,ss} = 2\Lambda_{mn}\,, \tag{1, 32}$$

---

[3] The choice of the coördinate conditions is, to a large extent arbitrary, and it might seem rather more natural to use the conditions

$$\eta^{\nu\mu}\gamma_{\alpha\nu|\mu} = 0$$

which are invariant under a Lorentz transformation. However, it turns out that the actual calculation of the field is simpler when we use the coördinate conditions given in the text and it is for this reason that we employ it in the general theory.

where Λ is related to L exactly as γ is to h:

$$\Lambda_{00} = \tfrac{1}{2}L_{00} + \tfrac{1}{2}L_{ll}\,, \tag{1, 33}$$

$$\Lambda_{0n} = L_{0n}\,, \tag{1, 34}$$

$$\Lambda_{mn} = L_{mn} - \tfrac{1}{2}\delta_{mn}L_{ll} + \tfrac{1}{2}\delta_{mn}L_{00}\,. \tag{1, 35}$$

These field equations, (1, 33), (1, 34), (1, 35) together with the coördinate conditions (1, 25), (1, 26) will form the basis of our further considerations.

2. **Fundamental Integral Properties of the Field.** Let us consider three functions A_n; $(n = 1, 2, 3)$. They need not be tensors. From these functions we may build the three further functions

$$(A_{n|s} - A_{s|n})_{|s}\,, \tag{2, 1}$$

which can be explicitly written as

$$\begin{gathered} \{(A_{1|2} - A_{2|1})_{|2} - (A_{3|1} - A_{1|3})_{|3}\}, \quad \{(A_{2|3} - A_{3|2})_{|3} - (A_{1|2} - A_{2|1})_{|1}\}, \\ \{(A_{3|1} - A_{1|3})_{|1} - (A_{2|3} - A_{3|2})_{|2}\}. \end{gathered} \tag{2, 2}$$

These three functions thus constitute the curl of the three functions

$$(A_{2|3} - A_{3|2}), \qquad (A_{3|1} - A_{1|3}), \qquad (A_{1|2} - A_{2|1}). \tag{2, 3}$$

Consider any surface S which does not pass through singularities of the field. Since (2, 1) is the curl of (2, 3), it follows from Stokes' theorem that the integral of the "normal"[4] component of (2, 1) over S is equal to the line integral of the tangential component of (2, 3) taken around the rim of S. If S is a closed surface its rim is of zero length so that the latter integral will vanish. We therefore have the theorem that, if S is any closed surface which does not pass through singularities of the field, then

$$\int (A_{n|s} - A_{s|n})_{|s} \cos(\mathbf{n}\cdot\mathbf{N})\, dS = 0, \tag{2, 4}$$

where $(\mathbf{n}\cdot\mathbf{N})$ denotes the "angle" between the direction of x^n and the "normal" to S, and the summation convention applies to the n. This theorem is valid whether S encloses singularities or not, and we shall now apply it to the present problem.

---

[4] Words like *normal*, *angle*, *sphere*, and so on are used here in a purely conventional sense to designate the corresponding functions of the coördinates x^m and equations which are implied by these terms in Euclidean geometry. The argument of this paragraph is independent of any particular metric, and we use the Euclidean nomenclature merely because it is apt and convenient.

From the coördinate conditions (1, 25), (1, 26) and the field equations (1, 31), (1, 32) we have

$$(2, 5) \qquad (\gamma_{0n|s} - \gamma_{0s|n})_{|s} = 2\Lambda_{0n} - \gamma_{00|0n} ,$$

$$(2, 6) \qquad (\gamma_{mn|s} - \gamma_{ms|n})_{|s} = 2\Lambda_{mn} .$$

We see that the left-hand sides of (2, 5), (2, 6) give four quantities of the form (2, 1), one coming from (2, 5) and three from (2, 6) for $m = 1, 2, 3$. It follows from (2, 4) that, if S is a surface which does not pass through singularities of the field,

$$(2,7) \qquad \int (\gamma_{00|0n} - 2\Lambda_{0n}) \cos (\mathbf{n}\cdot\mathbf{N})\, dS = 0,$$

$$(2,8) \qquad \int 2\Lambda_{mn} \cos (\mathbf{n}\cdot\mathbf{N})\, dS = 0.$$

From (2, 5), (2, 6) we see that, in those regions where there are no singularities,

$$(2, 9) \qquad (\gamma_{00|0n} - 2\Lambda_{0n})_{|n} = 0,$$

$$(2, 10) \qquad (2\Lambda_{mn})_{|n} = 0.$$

Therefore Gauss' theorem shows that if we take two closed surfaces S, S' such that no singularity lies on or between S and S', the integrals over S and S' give the same result. But the validity of the integral conditions for surfaces which enclose singularities, or more generally, which enclose regions where the field equations for empty space are not fulfilled, can only be shown by means of Stokes' theorem.

We are treating matter as a singularity in the field. Let us assume there are p bodies, each represented by a point singularity. The coördinates of each such singularity will be functions of the time alone. Since (2, 7), (2, 8) are valid for any S provided only that it does not pass through a singularity, we may choose p such surfaces, each enclosing only one of the p singularities, and thus obtain $4p$ distinct integral conditions. Each of these, being now independent of the shape of its S, will give a relation between the coördinates of the singularities and their time derivatives, and we shall see later that the integral conditions give, in fact, the *equations of motion* of the singularities. These equations are derived here from the field equations and coördinate conditions alone without any extraneous assumption.

If, instead of integrating around one singularity at a time, we integrate over a surface which contains all the singularities, we obtain the laws of conservation of energy and linear momentum for the whole system. These laws are, of course, merely consequences of the laws of motion for the individual particles but owing to many cancellations they take a comparatively simple form.

3. **The Method of Approximation.** The method of approximation which has been used up to now in the theory of relativity is as follows. We consider that in the equation

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu} \tag{3, 1}$$

the $h_{\mu\nu}$ depend continuously on a positive parameter λ in such a way that they vanish for $\lambda = 0$, so that for $\lambda = 0$ space-time becomes Galilean. We assume, therefore, that the $h_{\mu\nu}$ can be expanded in a power series[5] in λ:

$$h_{\mu\nu} = \sum_{l=1}^{\infty} \lambda^l \underset{l}{h}_{\mu\nu}. \tag{3, 2}$$

This expansion is introduced in the field equations which are then grouped according to the different powers of λ, taking the form

$$0 = R_{\mu\nu} = \sum_{l=1}^{\infty} \lambda^l \underset{l}{R}_{\mu\nu}. \tag{3, 3}$$

In order that a set of $h_{\mu\nu}$ depending on the parameter λ shall exist as a solution of the field equations it is necessary that each of the equations

$$\underset{l}{R}_{\mu\nu} = 0 \tag{3, 4}$$

shall be satisfied. The best known example of this method is its application to the first approximation.

We shall now show why this method of approximation is unsuitable for the treatment of quasi-stationary fields. If we introduce an energy tensor for the matter which produces the field we obtain for the first approximation, using imaginary time, the well-known equations

$$\gamma_{\mu\nu|\sigma\sigma} = -2T_{\mu\nu}, \tag{3, 5}$$

where the coördinate system is determined by the equations

$$\gamma_{\mu\sigma|\sigma} = 0. \tag{3, 6}$$

In the simplest case of incoherent matter (dust) producing the field we have

$$T_{\mu\nu} = \rho \frac{d\xi^\mu}{ds} \frac{d\xi^\nu}{ds}, \tag{3, 7}$$

where $d\xi^\mu/ds$ are the components of the velocity measured in terms of the proper time s. If we are concerned with a quasi-static situation, $d\xi^0/ds$ is of the order of magnitude of unity while the $d\xi^m/ds$ are relatively small. Thus in such a case we shall have

$$| T_{00} | \gg | T_{0n} | \gg | T_{mn} |, \tag{3, 8}$$

[5] In λ^l the l will always be an exponent, not a contravariant index!

and from the equations (3, 5) we must have correspondingly

$$|\gamma_{00}| \gg |\gamma_{0n}| \gg |\gamma_{mn}|. \tag{3, 9}$$

The usual method of approximation does not take this into account since it treats all the γ's as of the same order of magnitude although, in the quasi-static case, γ_{00} is very much larger than the other components of $\gamma_{\mu\nu}$. A really good method of approximation for the quasi-static case should make essential use of the relations (3, 9).

We are led to our present method of approximation most simply by considering the problem of constructing a method of approximation which is suitable for the solution of the approximate field equations (3, 5) for the quasi-static case. It turns out that the method of approximation to which we are led in this way is also suitable for the solution of the rigorous gravitational equations even when we are not dealing with quasi-static cases.

The first step is to give an explicit expression for the fact that the time derivative of a field quantity is small relative to the quantity itself and to its spatial derivatives. To do this we introduce an auxiliary time coördinate

$$\tau = \lambda x^0 \tag{3, 10}$$

and assume that every field quantity is a function of (τ, x^1, x^2, x^3) rather than of (x^0, x^1, x^2, x^3). If φ is such a quantity we now assume that φ, $\varphi_{|m}$ and $\partial\varphi/\partial\tau$ are of the same order of magnitude, so that $\varphi_{|0}$ is of the order of $\lambda\varphi$.

From this we conclude that if T_{00} in (3, 7) is of the order of magnitude of λ^q, then T_{0n} will be of the order of λ^{q+1} and T_{mn} of the order of λ^{q+2}.

Further, it follows from well-known considerations concerning the first approximation (the conservation of energy for the motion of a point) that γ_{00}, which is the potential energy of a unit mass, is of the same order of magnitude as the square of the velocity and is thus, in our present notation, of the order of λ^2. Hence we have the following orders of magnitude for the γ's:

$$\gamma_{00} \sim \lambda^2; \qquad \gamma_{0n} \sim \lambda^3; \qquad \gamma_{mn} \sim \lambda^4. \tag{3, 11}$$

If we expand the γ's as power series in λ we must therefore take the lowest powers of the expansions to be of the orders given in (3, 11). The fact that only second derivatives of the γ's with respect to the time enter the equation (3, 5) shows that the powers of λ in successive terms of the expansions of the γ's may differ by two.[6] We are thus led to the simple assumption that

$$\begin{aligned} \gamma_{00} &= \lambda^2 \underset{2}{\gamma_{00}} + \lambda^4 \underset{4}{\gamma_{00}} + \lambda^6 \underset{6}{\gamma_{00}} + \cdots, \\ \gamma_{0n} &= \lambda^3 \underset{3}{\gamma_{0n}} + \lambda^5 \underset{5}{\gamma_{0n}} + \cdots, \\ \gamma_{mn} &= \lambda^4 \underset{4}{\gamma_{mn}} + \lambda^6 \underset{6}{\gamma_{mn}} + \cdots. \end{aligned} \tag{3, 12}$$

[6] The omission of terms with λ^{2l+1} in γ_{00}, γ_{mn} and with λ^{2l} in γ_{0m} is possible and natural, but logically not strictly necessary. The addition of the omitted terms of (3, 12) could be made in such a way that it would correspond to an introduction of a retarded potential (outgoing wave). Such a procedure would however, be artificial though it would not influence the equations of motion derived in II, as will be shown elsewhere.

We cannot discuss the question of convergence in general, but it is of interest to show that the new method of approximation can give convergent results even where this would not at first be expected. We consider the case of the one-dimensional wave equation in its simplest form

$$f_{xx} - f_{tt} = 0. \tag{3, 13}$$

If, in accordance with the main idea of the new method of approximation, we write

$$\begin{aligned} f &= \underset{0}{f} + \lambda^2 \underset{2}{f} + \lambda^4 \underset{4}{f} + \cdots, \\ f_{xx} &= \underset{0}{f_{xx}} + \lambda^2 \underset{2}{f_{xx}} + \lambda^4 \underset{4}{f_{xx}} + \cdots, \\ f_{tt} &= \lambda^2 f_{\tau\tau} = \lambda^2 \underset{0}{f_{\tau\tau}} + \lambda^4 \underset{2}{f_{\tau\tau}} + \lambda^6 \underset{4}{f_{\tau\tau}} + \cdots, \end{aligned} \tag{3, 14}$$

we obtain from (3, 13) the successive equations

$$\underset{0}{f_{xx}} = 0, \tag{3, 15a}$$

$$\underset{2}{f_{xx}} - \underset{0}{f_{\tau\tau}} = 0, \tag{3, 15b}$$

$$\underset{4}{f_{xx}} - \underset{2}{f_{\tau\tau}} = 0, \tag{3, 15c}$$

From these equations we can find the general solution of the wave equation (3, 13) expressed as a power series in λ. For simplicity we shall consider only the case of a sinusoidal wave so that, out of the totality of solutions of (3, 15a),

$$\underset{0}{f} = A(\tau) + xB(\tau), \tag{3, 16}$$

we choose the particular solution[7]

$$\underset{0}{f} = \sin\tau \tag{3, 17a}$$

and at each subsequent stage of the procedure we ignore all arbitrary functions which may enter. From (3, 15b), (3, 15c), . . . , we thus find

$$\underset{2}{f} = -\frac{x^2}{2!}\sin\tau, \tag{3, 17b}$$

$$\underset{4}{f} = \frac{x^4}{4!}\sin\tau, \tag{3, 17c}$$

so that the solution takes the form

$$f = \sin\tau\left\{1 - \frac{(x\lambda)^2}{2!} + \frac{(x\lambda)^4}{4!} - \cdots\right\} = \cos(\lambda x)\sin\tau.$$

[7] The inclusion of the solution $\underset{0}{f} = x\sin\tau$ also leads to sinusoidal waves, as is easily seen.

On replacing τ by λt we have

$$f = \cos(\lambda x)\sin(\lambda t) \tag{3, 18}$$

which is an exact solution of (3, 13).

4. **Expansion Properties of Field Quantities.** We shall show in this section that there is a simple general rule concerning the types of expansion which will occur when we treat the gravitational equations by the present method of approximation. This rule is that

Any component having an odd number of zero suffixes will have only odd powers of λ in its expansion, while any component having an even number of such suffixes will involve only even powers of λ in its expansion.

The fundamental equations (3, 12) show that the $\gamma_{\mu\nu}$ conform to this rule. The relations (1, 19), (1, 20), (1, 21) between $\gamma_{\mu\nu}$ and $h_{\mu\nu}$ have inverse relations of precisely the same form with γ and h interchanged, as

$$h_{00} = \tfrac{1}{2}\gamma_{00} + \tfrac{1}{2}\gamma_{ll}, \tag{4, 1}$$

$$h_{0n} = \gamma_{0n}, \tag{4, 2}$$

$$h_{mn} = \gamma_{mn} - \tfrac{1}{2}\delta_{mn}\gamma_{ll} + \tfrac{1}{2}\delta_{mn}\gamma_{00}, \tag{4, 3}$$

and from (3, 12) it follows that the expansions for the h's in powers of λ are of the form

$$\begin{aligned} h_{00} &= \lambda^2 \underset{2}{h_{00}} + \lambda^4 \underset{4}{h_{00}} + \lambda^6 \underset{6}{h_{00}} + \cdots, \\ h_{0n} &= \lambda^3 \underset{3}{h_{0n}} + \lambda^5 \underset{5}{h_{0n}} + \cdots, \\ h_{mn} &= \lambda^2 \underset{2}{h_{mn}} + \lambda^4 \underset{4}{h_{mn}} + \lambda^6 \underset{6}{h_{mn}} + \cdots, \end{aligned} \tag{4, 4}$$

showing that the h's also conform to the general rule.

Further, since the $\eta_{\mu\nu}$ trivially conform because η_{0n} vanishes, it follows from (1, 4) that the $g_{\mu\nu}$ also conform.

We may write the relation

$$g_{\mu\nu}g^{\nu\sigma} = \delta^\sigma_\mu \tag{4, 5}$$

in the form

$$g_{\mu n}g^{n\sigma} + g_{\mu 0}g^{0\sigma} = \delta^\sigma_\mu. \tag{4, 6}$$

The two groups of terms on the left differ by an even number of zero suffixes so that, since the δ^σ_μ trivially conform to the general rule, we shall obtain enough equations at each approximation for finding the expansions of the $g^{\mu\nu}$ if we assume that the general rule is valid for these components too. However, the $g^{\mu\nu}$ are uniquely determined in terms of the $g_{\mu\nu}$ by (4, 5) so that the expansions according to the general rule will give the only solution and extraneous powers of λ will necessarily have zero coefficients. Thus the rule is applicable to the $g^{\mu\nu}$ and so, also, to the $h^{\mu\nu}$.

Let us consider next the Christoffel symbols of both kinds. We have

$$[\mu\nu, \sigma] = \tfrac{1}{2}(g_{\nu\sigma|\mu} + g_{\sigma\mu|\nu} - g_{\mu\nu|\sigma}), \tag{4, 7}$$

and since the operation "$_{|0}$" introduces a factor λ while the operations "$_{|m}$" leave the order of magnitude unchanged it is evident that the fact that the $g_{\mu\nu}$ obey the general rule implies that the $[\mu\nu, \sigma]$ do too.

The Christoffel symbols of the second kind are defined by

$$\left\{ \begin{matrix} \lambda \\ \mu\nu \end{matrix} \right\} = g^{\lambda\sigma}[\mu\nu, \sigma] \tag{4,8}$$

and since whenever we have a dummy suffix we shall have either no extra zero suffixes or two such suffixes entering any term in the implied summation, the fact that $g^{\lambda\sigma}$ and $[\mu\nu, \sigma]$ separately conform to the general rule shows that this is true also of the $\left\{ \begin{matrix} \lambda \\ \mu\nu \end{matrix} \right\}$.

In the course of the above considerations we have shown that neither the entry of dummy suffixes nor the operations "$_{|m}$", "$_{|0}$" disturbs the operation of the general rule. It follows that if, by the use of these operations alone, we form new quantities from quantities which conform to the rule these new quantities must also obey the general rule. This has already been exemplified by our discussion of the Christoffel symbols, and since all the quantities we shall have to consider, such as

$$R(= g^{\mu\nu}R_{\mu\nu}), \quad \Lambda_{\mu\nu}, \text{ etc.}$$

are new quantities of this type, we see that all the quantities with which we have to deal will have expansions in powers of λ whose general character is summed up in the statement at the head of this section.

5. Alternative Form of the Equations When Singularities Are Absent. In this section and the next we shall discuss the case where no singularities are present in the field. This case is, of course, trivial from the physical point of view since it corresponds to the complete absence of matter and, indeed, according to our method of approximation leads to the Galilean solution. Despite this, the discussion of this case will not be without value, for it will serve to exhibit the general mechanism of the theory and will form a convenient introduction to the later, more difficult discussion necessary when singularities are present.

Let us summarize some of the results we have obtained so far. The field has been subjected to the two restrictions

I The Gravitational Field Equations, and

II The Coördinate Conditions,

from which we have found

III The Surface Integral Conditions.

That is, if we take coördinate conditions

$$\gamma_{00|0} - \gamma_{0n|n} = 0, \tag{1, 25}$$

$$\gamma_{nm|n} = 0, \tag{1, 26}$$

the field equations take the form

$$\gamma_{00|ss} = 2\Lambda_{00}\,, \tag{1, 30}$$

$$\gamma_{0n|ss} = 2\Lambda_{0n}\,, \tag{1, 31}$$

$$\gamma_{mn|ss} = 2\Lambda_{mn}\,, \tag{1, 32}$$

and from these two groups of equations we obtain the surface integral conditions

$$\int (\gamma_{0c|0n} - 2\Lambda_{0n}) \cos(\mathbf{n}\cdot\mathbf{N})\, dS = 0, \tag{2, 7}$$

$$\int 2\Lambda_{mn} \cos(\mathbf{n}\cdot\mathbf{N})\, dS = 0, \tag{2, 8}$$

and also the results

$$(\gamma_{00|0n} - 2\Lambda_{0n})_{|n} = 2\Lambda_{00|0} - 2\Lambda_{0n|n} = 0, \tag{2, 9}$$

$$2\Lambda_{mn|n} = 0, \tag{2, 10}$$

which are essential for the validity of the surface integral conditions for arbitrary surfaces.

We shall now show that the following two sets of equations (5, 1), (5, 2) are equivalent when no singularities are present.

	(5, 1)		(5, 2)
(a)	$\gamma_{00\|ss} = 2\Lambda_{00}$,	(a)	$\gamma_{00\|ss} = 2\Lambda_{00}$,
(b)	$\gamma_{0n\|ss} = 2\Lambda_{0n}$,	(b)	$\gamma_{0n\|ss} = 2\Lambda_{0n}$,
(c)	$\gamma_{00\|0} - \gamma_{0n\|n} = 0$;	(c′)	$\Lambda_{00\|0} - \Lambda_{0n\|n} = 0$,
		(c″)	$\int (\gamma_{00\|0n} - 2\Lambda_{0n}) \cos(\mathbf{n}\cdot\mathbf{N})\, dS = 0$;
(d)	$\gamma_{mn\|ss} = 2\Lambda_{mn}$,	(d)	$\gamma_{mn\|ss} = 2\Lambda_{mn}$,
(e)	$\gamma_{mn\|n} = 0$,	(e′)	$\Lambda_{mn\|n} = 0$,
		(e″)	$\int 2\Lambda_{mn} \cos(\mathbf{n}\cdot\mathbf{N})\, dS = 0$.

In (5, 1) we have merely the field equations and coördinate conditions and we show essentially that the coördinate conditions may be replaced by the surface integral conditions[8] and the conditions (2, 9) (2, 10). The proof for the present

[8] When singularities are absent (5, 2c′), (5, 2c″) and also (5, 2e′), (5, 2e″) are equivalent equations, but we include them all here in order to facilitate comparison with the situation which arises when singularities are present.

case is trivial. For we have already shown that (5, 1) implies (5, 2) and the converse follows at once from the following considerations.

From (5, 2a) (5, 2b) and (5, 2c′) we have

$$(\gamma_{00|0} - \gamma_{0n|n})_{|ss} = 2\Lambda_{00|0} - 2\Lambda_{0n|n} = 0,$$

and since there are no singularities and the γ's must be zero at infinity this gives

$$\gamma_{00|0} - \gamma_{0n|n} = 0$$

which is (5, 1c). The proof for γ_{mn} is similar.

6. Splitting of the Equations When Singularities Are Absent. In the first section we gave a prescription for separating the terms of each of the field equations into two well-defined groups. In this section we shall discuss the splitting of the gravitational equations according to powers of λ and shall show why just this method of separation is implied by our method of approximation.

It is necessary first to introduce certain notations. Consider the quantity

$$h_{mn|0s}\,. \tag{6, 1}$$

When h_{mn} is expanded in powers of λ we write

$$h_{mn} = \lambda^2 \underset{2}{h_{mn}} + \lambda^4 \underset{4}{h_{mn}} + \cdots + \lambda^{2l} \underset{2l}{h_{mn}} + \cdots, \tag{6, 2}$$

where the numbers underneath the h's on the right serve the double purpose of distinguishing between the different functions h on the right and of showing with what power of λ each is associated in the expansion.

Now the fundamental assumption of our method of approximation requires that h_{mn} be a function of $(\lambda x^0, x^1, x^2, x^3)$ so that

$$h_{mn|s} = \frac{\partial h_{mn}}{\partial x^s}$$

but

$$h_{mn|0} = \frac{\partial h_{mn}}{\partial x^0} = \lambda\,\frac{\partial h_{mn}}{\partial \tau}.$$

In order to distinguish between ordinary differentiation with respect to (x^0, x^1, x^2, x^3) and ordinary differentiation with respect to (τ, x^1, x^2, x^3) we shall denote the latter by a comma followed by an appropriate suffix:

$$h_{mn|s} = \frac{\partial h_{mn}}{\partial x^s} = h_{mn,s}\,, \tag{6, 3}$$

$$h_{mn|0} = \frac{\partial h_{mn}}{\partial x^0} = \lambda\,\frac{\partial h_{mn}}{\partial \tau} = \lambda\, h_{mn,0}\,. \tag{6, 4}$$

Thus h_{mn}, $h_{mn,s}$ and $h_{mn,0}$ are all of the same order, but $h_{mn|0}$ belongs to a power of λ one higher.

With this convention we may write the expansion of (6, 1) in the form

$$(6,\ 5)\qquad h_{mn|0s} = \lambda h_{mn,0s} = \lambda^3 \underset{2}{h_{mn,0s}} + \lambda^5 \underset{4}{h_{mn,0s}} + \cdots + \lambda^{2l+1} \underset{2l}{h_{mn,0s}} + \cdots .$$

Now, however, the number underneath each h on the right no longer indicates directly the power of λ with which it is associated. We therefore write a 1 underneath each zero suffix following a comma for every h having a number underneath so that (6, 5) becomes

$$(6,\ 6)\qquad h_{mn|0s} = \lambda h_{mn,0s} = \lambda^3 \underset{2\quad 1}{h_{mn,0s}} + \lambda^5 \underset{4\quad 1}{h_{mn,0s}} + \cdots + \lambda^{2l+1} \underset{2l\quad 1}{h_{mn,0s}} + \cdots .$$

Thus now the sum of the numbers underneath each h gives the power of λ with which it is associated while the first of these numbers indicates the particular function h we are considering. This notation is then consistent with the natural notation for a product of h's.

We consider now what happens when we introduce the power series expansions for the h's in the equations (1, 27), (1, 28), (1, 29). On equating to zero the coefficients of the various powers of λ we shall obtain

$$(6,\ 7)\qquad \underset{2l}{h_{00,ss}} = 2\underset{2l}{L_{00}},$$

$$(6,\ 8)\qquad \underset{2l+1}{h_{0n,ss}} = 2\underset{2l+1}{L_{0n}},$$

$$(6,\ 9)\qquad \underset{2l}{h_{mn,ss}} = 2\underset{2l}{L_{mn}}.$$

The lowest h's are $\underset{2}{h_{00}}$, $\underset{3}{h_{0n}}$, and $\underset{2}{h_{mn}}$, and these will therefore be the quantities determined in the first approximation. They correspond to $l = 1$ in the scheme of (6, 7) (6, 8) (6, 9). Thus at any stage, say l, the quantities to be determined are $\underset{2l}{h_{00}}$, $\underset{2l+1}{h_{0n}}$, $\underset{2l}{h_{mn}}$, and the quantities already known from the solutions of the previous approximations are the h's having lower numbers underneath.

But if we look at the forms of the L's, as given in (1, 15), (1, 16), (1, 17) we see that at the stage l we have either quadratic terms or linear terms involving differentiations with respect to x^0. The quadratic terms can only involve h's of lower order than for l, and the linear terms may be written as

$$(6,\ 10)\qquad \underset{2l-1\ 1}{h_{0s,0s}} - \tfrac{1}{2}\underset{2l-2\ 11}{h_{ss,00}} \quad \text{in } \underset{2l}{L_{00}},$$

$$(6,\ 11)\qquad \text{none} \quad \text{in } \underset{2l+1}{L_{0n}},$$

$$(6,\ 12)\qquad \tfrac{1}{2}\underset{2l-2\ 11}{h_{mn,00}} - \tfrac{1}{2}\underset{2l-1\ 1}{h_{0m,0n}} - \tfrac{1}{2}\underset{2l-1\ 1}{h_{0n,0m}} \quad \text{in } \underset{2l}{L_{mn}}.$$

These are all known functions from the previous approximations. Thus the whole of $\underset{2l}{L_{00}}$, $\underset{2l+1}{L_{0n}}$, $\underset{2l}{L_{mn}}$ for given l are known from the solutions of the previous approximations. This is the reason for the particular method of separation of the field equations into two parts described in 1. When the separation is made in this manner and the power series expansions are inserted for the h's in (1, 27) (1, 28), (1, 29), for each power of λ the corresponding coefficients automatically group themselves into those quantities which enter for the first time with the approximation in question and those which are already known, at least in principle, from the previous approximations. These two groups correspond exactly to the left and right hand sides of (1, 27), (1, 28), (1, 29).

Before we can solve the approximation equations we must also split the coördinate conditions (1, 25), (1, 26), and the relations between the h's and γ's according to powers of λ. It turns out that we may take at each stage

$$\underset{2l}{\gamma_{00,ss}} = 2\underset{2l}{\Lambda_{00}}, \qquad \underset{2l+1}{\gamma_{0n,ss}} = 2\underset{2l+1}{\Lambda_{0n}}, \qquad \underset{2l}{\gamma_{00,\underset{1}{0}}} - \underset{2l+1}{\gamma_{0n,n}} = 0; \tag{6, 13}$$

$$\underset{2l}{\gamma_{mn,ss}} = 2\underset{2l}{\Lambda_{mn}}, \qquad \underset{2l}{\gamma_{mn,n}} = 0, \tag{6, 14}$$

where the Λ's are known because of the solutions of the previous approximations.

We may also split the alternative equations (5, 2) and use, instead, at each stage

$$\begin{gathered}\underset{2l}{\gamma_{00,ss}} = 2\underset{2l}{\Lambda_{00}}, \qquad \underset{2l+1}{\gamma_{0n,ss}} = 2\underset{2l+1}{\Lambda_{0n}}, \qquad \underset{2l}{\Lambda_{00,\underset{1}{0}}} - \underset{2l+1}{\Lambda_{0n,n}} = 0, \\ \int \left(\underset{2l}{\gamma_{00,\underset{1}{0}n}} - 2\underset{2l+1}{\Lambda_{0n}} \right) \cos(\mathbf{n}\cdot\mathbf{N})\, dS;\end{gathered} \tag{6, 15}$$

$$\underset{2l}{\gamma_{mn,ss}} = 2\underset{2l}{\Lambda_{mn}}, \qquad \underset{2l}{\Lambda_{mn,n}} = 0, \qquad \int 2\underset{2l}{\Lambda_{mn}} \cos(\mathbf{n}\cdot\mathbf{N})\, dS = 0. \tag{6, 16}$$

As in the case of the unsplit equations, the surface integral conditions are consequences of the others because of the absence of singularities, and the whole splitting actually presents no fundamental difficulties for this case.

7. The General Theory When Singularities Are Present.

The existence of singularities in the field introduces certain factors which make the theory developed for the regular case inadequate. For, although the equations of the field are undefined at the singularities, their validity in the regular region is sufficient to determine the motion of these singularities. The slightest alteration in the position of a singularity amounts to an arbitrarily large alteration for a point near enough to the singularity, and we are therefore not permitted to make use of approximate expressions for the equations of motion in the development of our method of approximation. This fact leads to a new difficulty, in the approximation method, which must be discussed more fully.

Let there be p particles producing the field. We may represent their positions at any time by means of their spatial coördinates $\overset{k}{\xi}{}^m(\tau)$, $k = 1, 2, \cdots, p$. At these points the field will be singular, but we may enclose each of the singularities within a small surface,[9] and then the region exterior to these p surfaces will be regular.

Although the equations (5, 1), (5, 2) are undefined at the singularities, they have meaning in the regular region and we shall show that they can still be regarded as in some sense equivalent. The discussion can be divided into two parts, one dealing with the (a), (b) and (c) equations, which involve the suffix zero, and the other with the remaining equations having only spatial suffixes. We consider the latter. The essential structure of the (d) and (e) equations is preserved if we omit the suffix m and write for the total field

(7, 1)		(7, 2)	
(d)	$\gamma_{n\|ss} = 2\Lambda_n,$	(d)	$\gamma_{n\|ss} = 2\Lambda_n,$
(e)	$\gamma_{n\|n} = 0,$	(e′)	$\Lambda_{n\|n} = 0,$
		(e″)	$\int 2\Lambda_n \cos(\mathbf{n}\cdot\mathbf{N})\, dS = 0.$

The proof that (7, 1) implies (7, 2) has already been given in essence in 2. To prove the converse we first obtain from (7, 2d)

$$\gamma_{n|nss} = 2\Lambda_{n|n}, \tag{7, 3}$$

this being valid outside the surfaces enclosing the singularities. To solve this we make an analytic continuation of the functions Λ_n into the interiors of these surfaces in such a way that $\Lambda_{n|n}$ is everywhere zero. This is certainly possible because of the validity of (7, 2e″). So (7, 3) now becomes

$$\gamma_{n|nss} = 0$$

which, being everywhere valid, has the unique solution

$$\gamma_{n|n} = 0$$

which is (7, 1e).

Thus we have shown that if we make an analytic continuation of Λ_n so that (7, 2e′) is valid everywhere, then (7, 1) and (7, 2) are equivalent outside the surfaces enclosing the singularities.

It is clear from the proof that the result will hold for any surfaces enclosing the singularities.

For the (a), (b) and (c) equations a similar proof can also be given. In this case it is necessary to make an analytic continuation of the quantities Λ_{00} and

[9] Throughout the argument we assume that we are dealing with the situation at some definite time τ, allowing time to flow again only after the argument is concluded.

Λ_{0n} in such a way that (5, 2c′) is valid everywhere, this being possible because of (5, 2c″). We omit the details of this part of the proof that (5, 1), (5, 2) may be considered as equivalent even where singularities are present and shall regard the proof as complete.

To show the difficulty brought in by the use of our approximation method let us now consider only the equations (7, 2d), (7, 2e′) omitting the surface integral (7, 2e″). These equations determine the field in each of the approximation steps if the motions of the singularities are prescribed. The motion of the particles is then arbitrary as, for example, in the electrodynamical problem and the field is determined in each of the approximation steps by the equations

$$\underset{2l}{\gamma_{n,ss}} = 2\underset{2l}{\Lambda_n}$$

$$\underset{2l}{\Lambda_{n,n}} = 0.$$

The contradiction is evident if we try to add to these equations the surface condition split according to our approximation method. We then have the additional equation

$$\int^k 2\underset{2l}{\Lambda_n} \cos(\mathbf{n}\cdot\mathbf{N}) = 0 \tag{7, 4}$$

where (k) on top of an integral sign means that the surface of integration encloses only the k-singularity. We have in (7, 4) an infinite set of equations containing the functions $\overset{k}{\xi}$ and their time derivatives. These equations cannot be satisfied by the arbitrarily given $\overset{k}{\xi}$ functions characterising the motion.

This also shows how the difficulty can be avoided. We have to consider instead of (5, 1) or (5, 2) a more general set of conditions governing the field which contains those equations as a particular case. Since it is the surface integral conditions which cause the trouble we remove (5, 2c″), (5, 2e″) from the set (5, 2) and consider the significance of what remains.

In making this generalisation we have, of course, gone beyond the gravitational equations to others which contain them as a special case, and we must now discuss what changes have been induced in (5, 1) by this generalisation.

Since the surface integrals are independent of the surfaces, their values will be functions of the time alone through the ξ's and their derivatives. There is therefore no loss of generality if we denote these integrals taken over the p surfaces enclosing the various singularities by $4\pi\overset{k}{c}_0(\tau)$, $4\pi\overset{k}{c}_m(\tau)$:

$$\begin{aligned} \frac{1}{4\pi}\int^k (\gamma_{00|0} - 2\Lambda_{0n}) \cos(\mathbf{n}\cdot\mathbf{N})\, dS &= \overset{k}{c}_0(\tau), \\ \frac{1}{4\pi}\int^k 2\Lambda_{mn} \cos(\mathbf{n}\cdot\mathbf{N})\, dS &= \overset{k}{c}_m(\tau). \end{aligned} \tag{7, 5}$$

With this notation we shall now prove that the following two sets of equations (7, 6), (7, 7) are equivalent in a certain sense which will be explained in the course of the proof:

(7, 6)		(7, 7)	
(a)	$\gamma_{00\|ss} = 2\Lambda_{00}$,	(a)	$\gamma_{00\|ss} = 2\Lambda_{00}$,
(b)	$\gamma_{0n\|ss} = 2\Lambda_{0n}$,	(b)	$\gamma_{0n\|ss} = 2\Lambda_{0n}$,
(c)	$\gamma_{00\|0} - \gamma_{0n\|n} = -\sum_{k=1}^{p} \left\{ \overset{k}{c}_0 / \overset{k}{r} \right\}$;	(c′)	$\Lambda_{00\|0} - \Lambda_{0n\|n} = 0$;
(d)	$\gamma_{mn\|ss} = 2\Lambda_{mn}$,	(d)	$\gamma_{mn\|ss} = 2\Lambda_{mn}$,
(e)	$\gamma_{mn\|n} = -\sum_{k=1}^{p} \left\{ \overset{k}{c}_m / \overset{k}{r} \right\}$,	(e′)	$\Lambda_{mn\|n} = 0$.

Here $\overset{k}{r}$ is the "distance" from x^n to the k-singularity:

$$\overset{k}{r} = \left[\left(x^s - \overset{k}{\xi}{}^s\right)\left(x^s - \overset{k}{\xi}{}^s\right) \right]^{\frac{1}{2}}. \tag{7, 8}$$

We may introduce the surfaces enclosing the singularities as before and these equations will certainly have meaning outside them. The proof of their equivalence can here too be broken up into two parts and we shall only prove the equivalence for the (d) and (e) parts. Omitting the suffix m as before, we have

(7, 9)		(7, 10)	
(d)	$\gamma_{n\|ss} = 2\Lambda_n$,	(d)	$\gamma_{n\|ss} = 2\Lambda_n$,
(e)	$\gamma_{n\|n} = -\sum_{k=1}^{p} \left\{ \overset{k}{c} / \overset{k}{r} \right\}$,	(e′)	$\Lambda_{n\|n} = 0$,

with the notation

$$\frac{1}{4\pi} \int^{k} 2\Lambda_n \cos(\mathbf{n}\cdot\mathbf{N})\, dS = \overset{k}{c}(\tau). \tag{7, 11}$$

We begin by proving that (7, 10) implies (7, 9) under certain conditions of continuation. It is no longer possible to make an analytic continuation of Λ_n in such a way that (e′) is everywhere satisfied since this would imply that the surface integral is necessarily zero. In fact, from (7, 11) we see, by Gauss' theorem, that the continuation must be such that

$$\frac{1}{4\pi} \int^{k} 2\Lambda_{n\|n}\, dv = \frac{1}{4\pi} \int^{k} 2\Lambda_n \cos(\mathbf{n}\cdot\mathbf{N})\, dS = \overset{k}{c}(\tau). \tag{7, 12}$$

It is simplest for our purposes to make the continuation in such a way that Λ_n and $\Lambda_{n|n}$ are continuous at the surfaces, and that $\Lambda_{n|n}$ has a constant sign inside each surface and satisfies (7, 12).

Such a continuation is possible for any surfaces surrounding the singularities and can be made in such a way that when these surfaces shrink to zero size the function $\Lambda_{n|n}$ goes over to a sum of Dirac δ-functions:

$$\Lambda_{n|n} \rightarrow 4\pi \sum_{k=1}^{p} \overset{k}{c}(\tau) \cdot \delta\left(x^1 - \overset{k}{\xi}{}^1\right) \delta\left(x^2 - \overset{k}{\xi}{}^2\right) \delta\left(x^3 - \overset{k}{\xi}{}^3\right). \tag{7, 13}$$

From (7, 10d) we have now

$$\gamma_{n|nss} = 2\Lambda_{n|n}$$

so that

$$\gamma_{n|n}(x) = -\frac{1}{4\pi} \int \frac{2\Lambda_{n|n}(x')}{r(x, x')} \, dv', \tag{7, 14}$$

where the integral is to be taken over the whole domain of x^n and $r(x, x')$ is the "distance" from x^n to x'^n:

$$r(x, x') = [(x^s - x'^s)(x^s - x'^s)]^{\frac{1}{2}}. \tag{7, 15}$$

Because of the validity of (7, 10e′) outside the surfaces we may write (7, 14) as

$$\gamma_{n|n}(x) = -\frac{1}{4\pi} \sum_{k=1}^{p} \int^k \frac{2\Lambda_{n|n}(x')}{r(x, x')} \, dv', \tag{7, 16}$$

the integrals being taken only over the interiors of the surfaces. On shrinking these surfaces we may regard $r(x, x')$ as constants over the various domains of integration and write

$$\gamma_{n|n}(x) = -\frac{1}{4\pi} \sum_{k=1}^{p} 1/\left(\overset{k}{r}(x)\right) \int^k 2\Lambda_{n|n} \, dv,$$

and by (7, 12) this is

$$\gamma_{n|n} = -\sum_{k=1}^{p} \left\{\overset{k}{c}(\tau)/\overset{k}{r}\right\}$$

which is (7, 9e).

We have therefore shown that with the analytic continuation used above the equations (7, 10) imply the equations (7, 9).

To prove the converse we form from (7, 9) the relation

$$(\gamma_{n|s} - \gamma_{n|n})_{|s} = 2\Lambda_n + \sum_{k=1}^{p} \left\{\overset{k}{c}(\tau)/\overset{k}{r}\right\}_{|n}. \tag{7, 17}$$

If we now form the surface integrals of the "normal" components of the two sides of this equation for each of the surfaces enclosing the singularities in turn, the left hand side will give zero, as explained in 2, and we shall have left

$$\int^k 2\Lambda_n \cos(\mathbf{n} \cdot \mathbf{N}) \, dS = -\int^k \left\{\overset{k}{c}/\overset{k}{r}\right\}_{|n} \cos(\mathbf{n} \cdot \mathbf{N}) \, dS$$
$$= -\overset{k}{c} \int^k \left\{1/\overset{k}{r}\right\}_{|n} \cos(\mathbf{n} \cdot \mathbf{N}) \, dS = 4\pi \overset{k}{c}(\tau)$$

which is (7.11).

That the validity of (7, 10e′) for the regular region is contained in (7, 9) is trivial, and the equivalence of (7, 6d, e) and (7, 7d, e′) is therefore proved.

The proof of the equivalence for the remaining equations of (7, 6), (7, 7) containing the suffix zero presents no essentially new problems and will be omitted.

The whole point of our elaborate procedure in writing all the equations of the field in two equivalent forms is now clear since the present generalisation from (5, 1) to (7, 6) could not be made in a convincing manner without the aid of the parallelism with (5, 2) and (7, 7).

Owing to the absence of the surface integral conditions in (7, 7), there is no longer any objection to the application of our method of approximation to the solution of this set of equations. The λ's will cause a splitting of the equations just as before, except that the surface integral conditions will be absent. However, at each stage we may write

(7, 18)
$$\frac{1}{4\pi}\int^{k}\Big(\underset{2l\ \ 1}{\gamma_{00,0n}} - \underset{2l+1}{2\Lambda_{0n}}\Big)\cos(\mathbf{n}\cdot\mathbf{N})\,dS = \underset{2l+1}{\overset{k}{c}_0(\tau)},$$
$$\frac{1}{4\pi}\int^{k} \underset{2l}{2\Lambda_{mn}}\cos(\mathbf{n}\cdot\mathbf{N})\,dS = \underset{2l}{\overset{k}{c}_m(\tau)},$$

and with this notation we have the result, in precisely the same manner as for (7, 6), (7, 7), that for each stage of the approximation the following sets of equations (7, 19), (7, 20) are equivalent:

	(7, 19)		(7, 20)
(a)	$\underset{2l}{\gamma_{00,ss}} = \underset{2l}{2\Lambda_{00}},$	(a)	$\underset{2l}{\gamma_{00,ss}} = \underset{2l}{2\Lambda_{00}},$
(b)	$\underset{2l+1}{\gamma_{0n,ss}} = \underset{2l+1}{2\Lambda_{0n}},$	(b)	$\underset{2l+1}{\gamma_{0n,ss}} = \underset{2l+1}{2\Lambda_{0n}},$
(c)	$\underset{2l\ \ 1}{\gamma_{00,0}} - \underset{2l+1}{\gamma_{0n,n}} = -\sum_{k=1}^{p}\Big\{\underset{2l+1}{\overset{k}{c}_0(\tau)}/\overset{k}{r}\Big\};$	(c′)	$\underset{2l\ \ 1}{\Lambda_{00,0}} - \underset{2l+1}{\Lambda_{0n,n}} = 0;$
(d)	$\underset{2l}{\gamma_{mn,ss}} = \underset{2l}{2\Lambda_{mn}},$	(d)	$\underset{2l}{\gamma_{mn,ss}} = \underset{2l}{2\Lambda_{mn}},$
(e)	$\underset{2l}{\gamma_{mn,n}} = -\sum_{k=1}^{p}\Big\{\underset{2l}{\overset{k}{c}_m(\tau)}/\overset{k}{r}\Big\},$	(e′)	$\underset{2l}{\Lambda_{mn,n}} = 0.$

In the actual solving of the equations it is simpler to work with the sets (7, 19) rather than with (7, 20). At each stage we have to solve equations of the type $\gamma_{,ss} = 2\Lambda$ and in order to make the whole solution unambiguous we must impose the conditions that the field shall be Galilean at infinity and that no harmonic functions of higher type than simple poles may be added to the partial solutions except insofar as their addition is forced by the coördinate

conditions (7, 19c), (7, 19e). Let us suppose we have been able to solve all the successive approximations. Then the quantities $\overset{k}{c}_0(\tau)$, $\overset{k}{c}_m(\tau)$ are given by the relations

$$\overset{k}{c}_0(\tau) = \sum_{l=1}^{\infty} \lambda^{2l+1} \underset{2l+1}{\overset{k}{c}_0}(\tau), \tag{7, 21}$$

$$\overset{k}{c}_m(\tau) = \sum_{l=1}^{\infty} \lambda^{2l} \underset{2l}{\overset{k}{c}_m}(\tau). \tag{7, 22}$$

Our solution will not in general be a solution of the gravitational equations since (7, 6), (7, 7) are more general than those equations. However, if we now put

$$\overset{k}{c}_0(\tau) = 0, \qquad \overset{k}{c}_m(\tau) = 0 \tag{7, 23}$$

we impose such conditions on the motions of the singularities that our solutions will indeed become the solutions of the gravitational equations we are actually interested in.

The differential equations (7, 23) for the ξ's are really independent of λ since they must be expressed in terms, not of the auxiliary time τ but of the true time x^0, and when this is done the λ's will be necessarily reabsorbed.

In practice, of course, it is impossible to carry the computation beyond the first few stages. Let us suppose, then, that we have been able to solve the successive approximations up to some stage $l = q$. In this case, if we put

$$\sum_{l=1}^{q} \lambda^{2l+1} \underset{2l+1}{\overset{k}{c}_0}(\tau) = 0, \qquad \sum_{l=1}^{q} \lambda^{2l} \underset{2l}{\overset{k}{c}_m}(\tau) = 0, \tag{7, 24}$$

we shall obtain solutions of the gravitational equations correct to terms of the order $(2q + 1)$, and the equations (7, 24) will give the approximate equations of motion up to this order.

8. **The Zero Coördinate Condition.** We show in this section that the solution of our equations can always be made in such a way that

$$\underset{21+1}{\overset{k}{c}_0}(\tau) = 0, \tag{8, 1}$$

thus showing that the conditions

$$\overset{k}{c}_0(\tau) = \sum_{l=1}^{\infty} \lambda^{2l+1} \underset{2l+1}{\overset{k}{c}_0}(\tau) \tag{7, 21}$$

place no restriction on the motion of the singularities. This result is of significance because the conditions (7, 22) are alone sufficient to describe the motion completely and any further condition, if not redundant, would cause an overdetermination of the motion.

We actually use (8, 1) as normalisation conditions for each stage of the approximation and they are essential for the uniqueness of the solution.

The significant equations for the present argument are

$$\underset{2l}{\gamma}_{00,ss} = 2\underset{2l}{\Lambda}_{00}, \tag{7, 19a}$$

$$\underset{2l+1}{\gamma}_{0n,ss} = 2\underset{2l+1}{\Lambda}_{0n}, \tag{7, 19b}$$

$$\underset{2l}{\gamma}_{00,\underset{1}{0}} - \underset{2l+1}{\gamma}_{0n,n} = -\sum_{k=1}^{p}\left\{\underset{2l+1}{\overset{k}{c}}{}_0(\tau)/\overset{k}{r}\right\}, \tag{7, 19c}$$

where the Λ's are known from the solutions of the previous approximations, and we shall suppose that we have a solution of these equations. If we introduce the quantities $\overset{k}{\Gamma}(\tau)$ by means of the equation

$$\underset{2l}{\overset{k}{\Gamma}}{}_{,\underset{1}{0}}(\tau) = \underset{2l+1}{\overset{k}{c}}{}_0(\tau), \tag{8, 2}$$

we may write (7, 19c) in the form

$$\begin{aligned} \underset{2l}{\gamma}_{00,\underset{1}{0}} - \underset{2l+1}{\gamma}_{0n,n} &= -\sum_{k=1}^{p}\left\{\left(\underset{2l}{\overset{k}{\Gamma}}/\overset{k}{r}\right)_{,\underset{1}{0}} - \underset{2l}{\overset{k}{\Gamma}}\left(1/\overset{k}{r}\right)_{,\underset{1}{0}}\right\} \\ &= -\sum_{k=1}^{p}\left\{\left(\underset{2l}{\overset{k}{\Gamma}}/\overset{k}{r}\right)_{,\underset{1}{0}} + \left(\underset{2l}{\overset{k}{\Gamma}}\,\underset{1}{\overset{k}{\dot{\xi}}}{}^{n}/\overset{k}{r}\right)_{,n}\right\}, \end{aligned} \tag{8, 3}$$

where $\dot{\xi} = \dfrac{d\xi}{d\tau}$.

From (7, 19a), (7, 19b) we see that $\underset{2l}{\gamma}_{00}$ and $\underset{2l+1}{\gamma}_{0n}$ are arbitrary to the extent of additive harmonic functions and we may therefore add simple poles to them to form the new quantities

$$\underset{2l}{\gamma'}_{00} = \underset{2l}{\gamma}_{00} + \sum_{k=1}^{p}\left\{\underset{2l}{\overset{k}{\Gamma}}/\overset{k}{r}\right\}, \tag{8, 4}$$

$$\underset{2l+1}{\gamma'}_{0n} = \underset{2l+1}{\gamma}_{0n} - \sum_{k=1}^{p}\left\{\underset{2l}{\overset{k}{\Gamma}}\,\overset{k}{\dot{\xi}}{}^{n}/\overset{k}{r}\right\}. \tag{8, 5}$$

These new γ's however, while still satisfying (7, 19a), (7, 19b) will be such that

$$\underset{2l}{\gamma'}_{00,\underset{1}{0}} - \underset{2l+1}{\gamma'}_{0n,n} = 0. \tag{8, 6}$$

Since the c_0's now vanish, the surface integrals will also be zero and thus the zero coördinate condition will not affect the motion. This theorem and our

previous results show us that the equations which must form the basis of the actual calculation of the field and the equations of motion of the singularities are:

$$\gamma_{\underset{2l}{00},ss} = 2\underset{2l}{\Lambda_{00}}, \tag{8, 7a}$$

$$\gamma_{\underset{2l+1}{0n},ss} = 2\underset{2l+1}{\Lambda_{0n}}, \tag{8, 7b}$$

$$\gamma_{\underset{2l}{00},\underset{1}{0}} - \gamma_{\underset{2l+1}{0n},n} = 0; \tag{8, 7c}$$

$$\gamma_{\underset{2l}{mn},ss} = 2\underset{2l}{\Lambda_{mn}}, \tag{8, 7d}$$

and

$$\gamma_{\underset{2l}{mn},n} = -\sum_{k=1}^{p} \left\{ \underset{2l}{\overset{k}{c}_m}(\tau)/\overset{k}{r} \right\}, \tag{8, 7e}$$

with

$$\underset{2l}{\overset{k}{c}_m}(\tau) = \frac{1}{4\pi} \int^{k} 2\underset{2l}{\Lambda_{mn}} \cos(\mathbf{n}\cdot\mathbf{N})\, dS. \tag{8, 8}$$

The approximate equations of motion for the stage $l = q$ are given by

$$\sum_{l=1}^{q} \lambda^{2l} \underset{2l}{\overset{k}{c}_m}(\tau) = 0. \tag{8, 9}$$

II. Application of the General Theory

Note. In the first part of this paper we developed the general theory of a new method for solving the equations of gravitation by successive approximation and for obtaining the equations of motion, in principle to any desired degree of accuracy. In the present part we deal with the actual application of this method, carrying the calculation to such a stage that the main deviation from the Newtonian laws of motion is determined.

Unfortunately, as the work proceeds, the calculations become more and more extensive involving a great amount of technical detail which can have no intrinsic interest. To give all these calculations explicitly here would be quite impracticable and we are obliged to confine ourselves to stressing the general ideas of the work and merely announcing the actual results. For the convenience of anyone who may be interested in the details of the calculation, however, the entire computation of this part of our paper has been deposited with the Institute for Advanced Study so as to be available for reference.[10]

9. **The Approximation $l = 1$.** The approximation $l = 0$ is trivial, leading to the Galilean case, and we proceed at once to the next approximation $l = 1$.

---

[10] c/o Secretary of the School of Mathematics, Institute for Advanced Study, Princeton, N. J. (U. S. A.).

Since the quantities $\underset{2}{\Lambda}_{00}$, $\underset{3}{\Lambda}_{0n}$, and, $\underset{2}{\Lambda}_{mn}$ are all zero and, as explained in 3, the $\underset{2}{\gamma}_{mn}$ are also zero, we have left from all the equations (8.7a), . . . , (8.7e), merely

$$\underset{2}{\gamma}_{00,ss} = 0, \tag{9.1}$$

$$\underset{3}{\gamma}_{0n,ss} = 0, \tag{9.2}$$

$$\underset{2}{\gamma}_{00,\underset{1}{0}} - \underset{3}{\gamma}_{0n,n} = 0. \tag{9.3}$$

The character of our whole solution will depend essentially upon the choice of the harmonic function we take as the solution of (9.1). We shall assume that the particles we are interested in have spherical symmetry and that the field is Galilean at infinity. In this case the solution of (9.1) is unique since each singularity in $\underset{2}{\gamma}_{00}$ must now, by (9.1) be a simple pole. We therefore have for $\underset{2}{\gamma}_{00}$ the solution

$$\underset{2}{\gamma}_{00} = 2\varphi, \qquad \varphi = \sum_{k=1}^{p}\left\{-2\overset{k}{m}/\overset{k}{r}\right\}, \qquad \overset{k}{r} = \left[\left(x^s - \overset{k}{\xi}{}^s\right)\left(x^s - \overset{k}{\xi}{}^s\right)\right]^{\frac{1}{2}}, \tag{9.4}$$

where the p quantities $\overset{k}{m}$ are independent of the spatial coördinates x^s, and can depend at most only on the time.

From (9.2) we see that $\underset{3}{\gamma}_{0n}$ is also a harmonic function, and to determine it more exactly we must use the coördinate (9.3). From (9.3), (9.4) we have

$$\begin{aligned}\underset{3}{\gamma}_{0n,n} = \underset{2}{\gamma}_{00,\underset{1}{0}} &= \sum_{k=1}^{p}\left(-4\overset{k}{m}/\overset{k}{r}\right)_{,0} \\ &= \sum_{k=1}^{p}\left\{\left(4\overset{k}{m}/\overset{k}{r}\right)\dot{\overset{k}{\xi}}{}^n\right\}_{,n} - \sum_{k=1}^{p}\left(-4\dot{\overset{k}{m}}/\overset{k}{r}\right).\end{aligned}$$

This equation can be solved without introducing new singularities only if $\dot{\overset{k}{m}} = 0$. In other words, the quantities $\overset{k}{m}$, which actually measure the masses of the point singularities, are necessarily constants. It is now evident that, under our general restricting conditions, $\underset{3}{\gamma}_{0n}$ is uniquely determined:

$$\underset{3}{\gamma}_{0n} = \sum_{k=1}^{p}\left\{\left(4\overset{k}{m}/\overset{k}{r}\right)\dot{\overset{k}{\xi}}{}^n\right\}. \tag{9.5}$$

In all that follows we shall limit our considerations to the case of only two particles. This places no essential restriction on the results as far as the end of 15, their generalisation to p particles being trivial, and it permits a useful simplification of the rather inconvenient notation used for the general case.

For the case of two particles we shall write:

$$\text{(9.6)}\qquad \begin{aligned} &\text{(a)}\quad -2\overset{1}{m}/\overset{1}{r} = \psi, \qquad -2\overset{2}{m}/\overset{2}{r} = \chi;\\ &\text{(b)}\quad \varphi = \psi + \chi;\\ &\text{(c)}\quad \overset{1}{\xi}{}^{s} = \eta^{s}, \qquad \overset{2}{\xi}{}^{s} = \zeta^{s}. \end{aligned}$$

Our results (9.4), (9.5) may thus be written in the form

$$\text{(9.7)}\qquad \begin{aligned} &\text{(a)}\quad \underset{2}{\gamma_{00}} = 2\varphi = 2\psi + 2\chi,\\ &\text{(b)}\quad \underset{3}{\gamma_{0n}} = -2\psi\dot{\eta}^{n} - 2\chi\dot{\zeta}^{n}. \end{aligned}$$

From (1.18) we now also have

$$\text{(9.8)}\qquad \begin{aligned} &\text{(a)}\quad \underset{2}{h_{00}} = \varphi = \psi + \chi,\\ &\text{(b)}\quad \underset{3}{h_{0n}} = -2\psi\dot{\eta}^{n} - 2\chi\dot{\zeta}^{n},\\ &\text{(c)}\quad \underset{2}{h_{mn}} = \delta_{mn}\varphi = \delta_{mn}(\psi + \chi). \end{aligned}$$

This shows that the approximation $l = 1$ has a Newtonian character but, owing to the vanishing of $\underset{2}{\overset{k}{c}}{}_{m}$, places no restriction on the motion.

10. **Calculation of the Λ's for $l = 2$.** The first step in the calculation of the Λ's for $l = 2$ is the determination of the $h_{\mu\nu}$.

Using the method explained in 4, we can calculate the expansions of the $h^{\mu\nu}$ to any desired degree of approximation. We find, for $l = 1$,

$$\text{(10.1)}\qquad \underset{2}{h^{00}} = -\underset{2}{h_{00}} = -\varphi,$$

$$\text{(10.2)}\qquad \underset{3}{h^{0n}} = \underset{3}{h_{0n}} = \underset{3}{\gamma_{0n}},$$

$$\text{(10.3)}\qquad \underset{2}{h^{mn}} = -\underset{2}{h_{mn}} = -\delta_{mn}\varphi.$$

We next have to calculate for $l = 2$ the quantities $2L_{\mu\nu}$ defined in (1.15), (1.16), (1.17).

In $2\underset{4}{L_{00}}$ the linear terms give

$$\varphi_{,00}\,.$$

Of the non-linear terms, only three can give a contribution. They are

$$-2\left\{\underset{2}{h^{sr}}\left[\underset{2}{00, r}\right]\right\}_{,s} = -\varphi_{,s}\varphi_{,s} \quad (\text{since } \varphi_{,ss} = 0),$$

$$-2\left[\underset{2}{00, s}\right]\left[\underset{2}{0s, 0}\right] = \tfrac{1}{2}\varphi_{,s}\varphi_{,s}\,,$$

$$-2\left[\underset{2}{00, r}\right]\left[\underset{2}{rs, s}\right] = \tfrac{3}{2}\varphi_{,s}\varphi_{,s}\,,$$

where $[rs, p]$ are Christoffel symbols. Thus

$$2\underset{4}{L}_{00} = \varphi_{,00} + \varphi_{,s}\varphi_{,s}\,. \tag{10.4}$$

Similar but rather more tiresome calculations lead to the further results

$$2\underset{5}{L}_{0n} = \varphi_{,s}\underset{3}{h}_{0s,n} - \varphi_{,sn}\underset{3}{h}_{0s} - 3\varphi_{,0}\varphi_{,n}\,, \tag{10.5}$$

$$2\underset{4}{L}_{mn} = -\underset{3}{h}_{0m,0n} - \underset{3}{h}_{0n,0m} + \delta_{mn}\varphi_{,00} - 2\varphi\varphi_{,mn} - \varphi_{,m}\varphi_{,n} - \delta_{mn}\varphi_{,s}\varphi_{,s}\,. \tag{10.6}$$

Therefore, by (1.30), $\cdots$, (1.35), we have

$$\begin{aligned}
&\text{(a)} \quad \underset{4}{\gamma}_{00,ss} = 2\underset{4}{\Lambda}_{00} = -\tfrac{3}{2}\varphi_{,s}\varphi_{,s}\,,\\
&\text{(b)} \quad \underset{5}{\gamma}_{0n,ss} = 2\underset{5}{\Lambda}_{0n} = \varphi_{,s}\underset{3}{\gamma}_{0s,n} - \varphi_{,sn}\underset{3}{\gamma}_{0s} - 3\varphi_{,0}\varphi_{,n}\,,\\
&\text{(c)} \quad \underset{4}{\gamma}_{mn,ss} = 2\underset{4}{\Lambda}_{mn} = -\underset{3}{\gamma}_{0m,0n} - \underset{3}{\gamma}_{0n,0m} + 2\delta_{mn}\varphi_{,00} - 2\varphi\varphi_{,mn} - \varphi_{,m}\varphi_{,n}\\
&\qquad\qquad + \tfrac{3}{2}\delta_{mn}\varphi_{,s}\varphi_{,s}\,.
\end{aligned} \tag{10.7}$$

As explained in 7, 8, these equations (10.7), together with the corresponding coördinate conditions

$$\begin{aligned}
&\text{(a)} \quad \underset{4}{\gamma}_{00,\underset{1}{0}} - \underset{5}{\gamma}_{0n,n} = 0,\\
&\text{(b)} \quad \underset{4}{\gamma}_{mn,n} = -\Big(\underset{4}{\overset{1}{c}}_m/\overset{1}{r}\Big) - \Big(\underset{4}{\overset{2}{c}}_m/\overset{2}{r}\Big),
\end{aligned} \tag{10.8}$$

are the equations which determine the field in the next approximation.

11. **The Newtonian Equations of Motion.** We must now evaluate the surface integrals

$$\underset{4}{\overset{k}{c}}_m(\tau) = \frac{1}{4\pi}\int^k 2\underset{4}{\Lambda}_{mn}\cos(\mathbf{n}\cdot\mathbf{N})\,dS, \qquad k = 1, 2. \tag{11.1}$$

According to the general theory of part I, these integrals will be independent of the particular shapes of the surfaces of integration since the divergences of their integrands must vanish on a consequence of the field equations belonging to the previous approximation. We shall show here by actual calculation that this is the case with the $2\underset{4}{\Lambda}_{mn}$ given in (10.7c).

Since φ and $\underset{3}{\gamma}_{0n}$ are harmonic functions, we have

$$2\underset{4}{\Lambda}_{mn,n} = -\underset{3}{\gamma}_{0n,0mn} + 2\varphi_{,00m}$$

which is zero, as can easily be seen from (9.3) and (9.7a).

In the actual calculation of the surface integrals we evaluate the separate contributions of the different terms in $2\underset{4}{\Lambda}_{mn}$. Since the value of a whole

integral is independent of the shape of its surface of integration, by taking this surface to be of finite size and always a finite distance from its singularity, we see that the whole integral cannot be infinite. Now the individual terms of $2\underset{4}{\Lambda}_{mn}$ have not the property that their divergences vanish, and so we must fix the surfaces of integration quite definitely before we begin the calculations. It is most convenient to take definite, infinitesimally small spheres whose centers are at the singularities, but in this case infinities of the types

$$\lim_{r\to 0} \text{const.}/r^n, \qquad n \text{ a positive integer,}$$

can occur in the values of the partial integrals. Since these must cancel, however, in the final result, we may merely ignore them throughout the calculation of the surface integrals.

We shall consider the integral taken around the first singularity. Owing to the infinitesimal size of the surface of integration, the only terms which can give results different from zero or infinity are those of the order of $\left(1/\overset{1}{r}{}^2\right)$.

The first term in $2\underset{4}{\Lambda}_{mn}$ is $-\underset{3}{\gamma}_{0m,0n}$, and, by (8.7b), this may be written as

$$-\underset{3}{\gamma}_{0m,0n} = -2\psi_{,ns}\dot{\eta}^m\dot{\eta}^s + 2\psi_{,n}\ddot{\eta}^m - 2\chi_{,ns}\dot{\zeta}^m\dot{\zeta}^s + 2\chi_{,n}\ddot{\zeta}^m.$$

The only term we need consider is the second, and so we have

$$\frac{1}{4\pi}\int^1 \left(-\underset{3}{\gamma}_{0m,0n}\right)\cos(\mathbf{n}\cdot\mathbf{N})\,dS = \frac{1}{4\pi}\int^1 2\psi_{,n}\ddot{\eta}^m\cos(\mathbf{n}\cdot\mathbf{N})\,dS$$

$$(11.2) \qquad = \left(4\overset{1}{m}\,\ddot{\eta}^m\right)\frac{1}{4\pi}\int^1\left\{(x^n - \eta^n)(x^n - \eta^n)/\overset{1}{r}{}^4\right\}dS$$

$$= \left(4\overset{1}{m}\,\ddot{\eta}^m\right)\frac{1}{4\pi}\int^1\left\{1/\overset{1}{r}{}^2\right\}dS = 4\overset{1}{m}\ddot{\eta}^m.$$

In a similar manner we find that

$$(11.3) \qquad \frac{1}{4\pi}\int^1\left(-\underset{3}{\gamma}_{0n,0m}\right)\cos(\mathbf{n}\cdot\mathbf{N})\,dS = \tfrac{4}{3}\overset{1}{m}\ddot{\eta}^m.$$

The fourth term, $(-2\varphi, \varphi_{,mn})$, requires slightly different treatment. The only part that can be of interest is

$$-2\psi_{,mn}\,\chi$$

and in order to evaluate the corresponding contribution to the surface integral we must expand χ as a power series in the neighborhood of the first singularity, writing

$$(11.4) \qquad \chi = \tilde{\chi} + (x^s - \eta^s)\tilde{\chi}_{,s} + \cdots,$$

where

$$(11.5) \qquad \tilde{\chi} = \chi(\eta^n), \qquad \tilde{\chi}_{,s} = \chi_{,s}(\eta^n), \text{ etc.}$$

Introducing this expansion for χ we see that the only term in the integrand which can give a finite result is

$$-2\psi_{,mn}(x^s - \eta^s)\tilde{\chi}_{,s} \tag{11.6}$$

The determination of the surface integral of this term depends on the calculation of

$$\int^1 (x^s - \eta^s)\, \psi_{,mn} \cos(\mathbf{n}\cdot\mathbf{N})\, dS.$$

We have

$$(x^s - \eta^s)\, \psi_{,mn} \cos(\mathbf{n}/\mathbf{N})$$

$$= 2\overset{1}{m}(x^s - \eta^s)\left\{-3(x^m - \eta^m)(x^n - \eta^n)/\overset{1}{r}^5 + \delta_{mn}/\overset{1}{r}^3\right\}(x^n - \eta^n)/\overset{1}{r}$$

$$= -4\overset{1}{m}(x^s - \eta^s)(x^m - \eta^m)/\overset{1}{r}^4.$$

Therefore

$$\frac{1}{4\pi}\int (x^s - \eta^s)\, \psi_{,mn} \cos(\mathbf{n}\cdot\mathbf{N})\, dS = -\frac{4\overset{1}{m}}{4\pi}\int (x^s - \eta^s)(x^m - \eta^m)/\overset{1}{r}^4\, dS$$

$$= -\frac{4\overset{1}{m}}{3}\delta_{ms}, \tag{11.7}$$

and so the surface integral of the term (11.6) is

$$+\frac{8\overset{1}{m}}{3}\tilde{\chi}_{,m}, \tag{11.8}$$

which is thus also the value of the surface integral for the whole of the term $(-\varphi\varphi_{,mn})$.

In a somewhat similar way we obtain, for the surface integrals of the remaining terms, the values

$$\left\{\begin{aligned} 2\delta_{mn}\varphi_{,00} &\rightarrow -\frac{4\overset{1}{m}}{3}\ddot{\eta}^m, \\ -\varphi_{,m}\varphi_{,n} &\rightarrow -\frac{8\overset{1}{m}}{3}\tilde{\chi}_{,m}, \\ \tfrac{3}{2}\delta_{mn}\varphi_{,s}\varphi_{,s} &\rightarrow 2\overset{1}{m}\tilde{\chi}_{,m}. \end{aligned}\right. \tag{11.9}$$

Hence we have

$$\underset{4}{\overset{1}{c}}_m(\tau) = \frac{1}{4\pi}\int^1 2\underset{4}{\Lambda}_{mn} \cos(\mathbf{n}\cdot\mathbf{N})\, dS = 4\overset{1}{m}\{\ddot{\eta}^m + \tfrac{1}{2}\tilde{\chi}_{,m}\}. \tag{11.10}$$

Let us assume for the moment that we are not going any further with the approximation. In this case our approximate equations of motion would be of the form

$$\lambda^4\{\ddot{\eta}^m + \tfrac{1}{2}\tilde{\chi}_{,m}\} = 0 \tag{11.11}$$

for each particle. It is of interest to note that this form of the equations of motion is actually independent of the variables x^s. For we have, by (11.5), (9.6),

$$\tilde{\chi}_{,s} = \chi_{,s}(\eta^n), \qquad \chi = -2\overset{2}{m}/\overset{2}{r}. \tag{11.12}$$

For our present argument we may take χ as any function of $\overset{2}{r}$. Equations (11.12) show that to form $\tilde{\chi}_{,s}$ we must first differentiate χ with respect to x^s and then replace x^s by η^s. But the result will be the same if we first replace x^s by η^s and later differentiate with respect either to η^s or to $(-\zeta^s)$. Thus

$$\tilde{\chi}_{,s} = \frac{\partial\chi(r)}{\partial\eta^s} = -\frac{\partial\chi(r)}{\partial\zeta^s}, \tag{11.13}$$

where r denotes the "distance" between η^s and ζ^s:

$$r = [(\eta^s - \zeta^s)(\eta^s - \zeta^s)]^{\frac{1}{2}}. \tag{11.14}$$

We can therefore think of our equations of motion as involving the differentiation of functions depending only on the positions of the singularities, as is characteristic of theories based on the concept of action at a distance.

Writing (11.11) more explicitly in vector notation as

$$\overset{1}{m}\ddot{\eta} = \nabla\left(\overset{1}{m}\overset{2}{m}/r\right), \tag{11.15}$$

we see that (11.11) gives precisely the Newtonian law of motion.[11]

We have therefore obtained the Newtonian equations of motion from the field equations alone, without extra assumption such as was hitherto believed to be necessary and was supplied by the law of geodetic lines, or by a special choice of an energy impulse tensor.

From the above derivation of the Newtonian equations of motion, the general mechanism becomes apparent by which the Lorentz equations for the motion of electric particles can be obtained. In this case we have to consider the gravitational equations in which the Maxwell energy-momentum tensor appears on the right, and also the Maxwell field equations, and treat the whole set of equations by our approximation method. It is necessary, now, to give each singularity an electric charge e in addition to its mass m. We may safely ignore the contribution arising from the products of gravitational potentials in

[11] Equation (11.11) and (11.15) are written in terms of the auxiliary time and the auxiliary masses. We shall return to this point in 17.

the new field equations. For this omission has the effect of destroying the second term of (11.11), while the inclusion of the Maxwell tensor leads to the appearance, on the right of (11.11), of a corresponding surface integral giving the electrostatic force acting on the particle. In the next approximation we obtain the full Lorentz force together with the relativistic correction to the mass.

So long as we are dealing with singularities, we have no basis within the theory for excluding negative masses; in other words, for excluding gravitational repulsions between particles. If, however, we decide always to take mass positive, then the sign with which the Maxwell energy-momentum tensor enters the field equations determines whether like charges shall attract or repel each other. This also reveals the limitations of any theory based upon the existence of singularities.

12. **Normalisation of $\underset{4}{\gamma_{00}}$.** The value of $\underset{4}{\gamma_{00}}$ determined from (10.7a) is arbitrary to within an added harmonic function, and this function is to be determined from the relations (8.4), (8.2), together with our basic requirement that higher harmonic functions than simple poles are, as far as possible, to be avoided.

From (10.7a) and the fact that φ is harmonic, we have at once

$$\underset{4}{\gamma_{00}} = -\tfrac{3}{4}\varphi\varphi + \alpha_{00}\psi + \beta_{00}\chi, \tag{12.1}$$

where we have written the additive functions of (8.4) in a different form more in accordance with our present notation, α_{00}, β_{00} being functions of τ alone through η and ζ and their derivatives. The quantities α_{00}, β_{00} can be determined from the condition that

$$\frac{1}{4\pi}\int\left\{\underset{4}{\gamma}_{00,0n} - 2\underset{5}{\Lambda}_{0n}\right\}\cos(\mathbf{n}\cdot\mathbf{N})\,dS = 0. \tag{12.2}$$

The value of α_{00} is found by taking this integral over a small sphere having its center at the first singularity, and from calculations similar to those of 3, we find, after making use of the equations of motion of the first order:

$$\alpha_{00} = \{\dot{\eta}^s\dot{\eta}^s + \tfrac{1}{2}\tilde{\chi}\}. \tag{12.3}$$

Similarly, by integrating over a small sphere around the second singularity, we find

$$\beta_{00} = \{\dot{\zeta}^s\dot{\zeta}^s + \tfrac{1}{2}\tilde{\psi}\}, \tag{12.4}$$

where

$$\begin{aligned}\tilde{\chi} &= \chi(\eta^s),\\ \tilde{\psi} &= \psi(\zeta^s).\end{aligned} \tag{12.5}$$

These results show clearly the physical significance of the particular normalisation required by the conditions (8.4), (8.2). For we now have

$$(12.6) \quad \lambda^2 \underset{2}{\gamma_{00}} + \lambda^4 \underset{4}{\gamma_{00}} = \lambda^2 \{(1 + \tfrac{1}{2}\lambda^2 \alpha_{00})\, 2\psi + (1 + \tfrac{1}{2}\lambda^2 \beta_{00})\, 2\chi - \tfrac{3}{4}\lambda^2 \varphi\varphi\},$$

and we see from (12.3), (12.4) that $\overset{1}{m}(1 + \tfrac{1}{2}\lambda^2 \alpha_{00})$, $m(1 + \tfrac{1}{2}\lambda^2 \beta_{00})$ involve the first relativistic corrections to the masses.

The calculations up to this stage correspond to those of Droste, De Sitter, and Levi-Civita, cited in the introduction.

13. **Solution of the Field Equations for $l = 2$.** Since our ultimate aim is to determine the equations of motion up to the next approximation, we are interested only in those expressions which give a contribution to the corresponding surface integrals. We shall state dogmatically what is needed for these calculations for the justification of our statement can not be given without exposing the details of our actual calculation.

1. The calculation of $\underset{4}{\gamma_{mn}}$ and $\underset{5}{\gamma_{0m}}$ in the neighborhood of the singularities. We do not need to care in $\underset{4}{\gamma_{mn}}$ about those terms which do not go to infinity if $\overset{1}{r} \rightarrow 0$.

2. The calculation of $\underset{4}{\gamma_{rr}}$ in the whole space.

The expression $2\underset{4}{\Lambda_{mn}}$ in (10, 7) can be divided into two parts, one containing the linear terms together with all other terms not involving interactions between the two particles, and the other containing all the interaction terms. We denote these two groups of terms respectively by X_{mn} and Y_{mn}. The integration of the equations

$$(13.1) \qquad \underset{4}{\gamma'_{mn,ss}} = X_{mn}$$

presents no difficulties, but the equations

$$(13.2) \qquad \underset{4}{\gamma''_{mn,ss}} = Y_{mn}$$

cannot, apparently, be integrated in an elementary manner and we are obliged to introduce a simplification. Since we need to know the values of $\underset{4}{\gamma_{mn}}$ mainly in order to evaluate the surface integrals $\underset{6}{c_m}$ about, say, the first particle, we may introduce power series expansions for χ in the neighborhood of this point and so obtain a solution for $\underset{4}{\gamma_{mn}}$ which is also in the form of such an expansion.

We find, actually, from (13.1), (13.2), the following expressions for $\underset{4}{\gamma}'_{mn}$ and $\underset{4}{\gamma}''_{mn}$:

$$\begin{aligned}\underset{4}{\gamma}'_{mn} = {} & \{\psi[(x^n - \eta^n)\dot\eta^m + (x^m - \eta^m)\dot\eta^n - \delta_{mn}(x^s - \eta^s)\dot\eta^s]\}_{,0} \\ & + \{\chi[(x^n - \zeta^n)\dot\zeta^m + (x^m - \zeta^m)\dot\zeta^n - \delta_{mn}(x^s - \zeta^s)\dot\zeta^s]\}_{,0} \\ & + \tfrac{7}{4}\overset{1}{r}{}^2\psi_{,m}\psi_{,n} + \tfrac{7}{4}\overset{2}{r}{}^2\chi_{,m}\chi_{,n},\end{aligned} \tag{13.3}$$

and

$$\underset{4}{\gamma}''_{mn} = -\psi_{,m}(x^n - \eta^n)\tilde\chi, \tag{13.4}$$

where we have included in (13.4) only those terms which ultimately have importance for the evaluation of the surface integrals $\underset{6}{\overset{1}{c}}_m$.

The value of $\underset{4}{\gamma}_{mn}$ is given by

$$\underset{4}{\gamma}_{mn} = \underset{4}{\gamma}'_{mn} + \underset{4}{\gamma}''_{mn} + \alpha_{mn}\psi, \tag{13.5}$$

where α_{mn} is a function of time to be determined from the coördinate conditions

In a similar way, we may calculate the values of $\underset{5}{\gamma}_{0n}$ in two parts. We find, on including only relevant terms for the surface integrals $\underset{6}{\overset{1}{c}}_m$, in the integrands of which $\underset{5}{\gamma}_{0n}$ enters only linearly,

$$\underset{5}{\gamma}'_{0m} = -\tfrac{7}{4}\overset{1}{r}{}^2\psi_{,m}\psi_{,s}\dot\eta^s + \tfrac{3}{4}\psi\psi\dot\eta^m, \tag{13.6}$$

$$\begin{aligned}\underset{5}{\gamma}''_{0n} = {} & -\tfrac{3}{2}(x^s - \eta^s)\psi\tilde\chi_{,m}\dot\zeta^s - (x^m - \eta^m)\psi\tilde\chi_{,s}\dot\eta^s \\ & + \tfrac{1}{2}\psi_{,m}(x^s - \eta^s)(x^l - \eta^l)\tilde\chi_{,l}\dot\zeta^s + (x^m - \eta^m)\psi\tilde\chi_{,s}\dot\zeta^s \\ & + \tfrac{1}{2}(x^s - \eta^s)\psi\tilde\chi_{,s}\dot\zeta^m + \tfrac{3}{2}(x^s - \eta^s)\psi\tilde\chi_{,m}\dot\eta^s \\ & + (x^s - \eta^s)\psi_{,m}\tilde\chi\dot\zeta^s.\end{aligned} \tag{13.7}$$

The value of $\underset{5}{\gamma}_{0n}$ is given by

$$\underset{5}{\gamma}_{0n} = \underset{5}{\gamma}'_{0n} + \underset{5}{\gamma}''_{0n} + \alpha_{0n}\psi, \tag{13.8}$$

where α_{0n} is a function of time to be determined from the normalisation condition.

It remains only to calculate $\underset{4}{\gamma}_{rr}$ in the whole space. From (10.7c) we have

$$\underset{4}{\gamma}_{rr,ss} = 2\varphi_{,00} + \tfrac{7}{2}\varphi_{,s}\varphi_{,s}, \tag{13.9}$$

therefore

$$\underset{4}{\gamma}_{rr} = -2\overset{1}{m}\overset{1}{r}_{,00} - 2\overset{2}{m}\overset{2}{r}_{,00} + \tfrac{7}{4}\varphi^2 + \alpha\psi + \beta\chi \tag{13.10}$$

where α and β are functions of time to be determined in such a way that $\underset{4}{\gamma}_{rr}$ in (13.10) would agree with $\underset{4}{\gamma}_{rr}$ determined from (13.5) near to the singularities.

14. **Determination of α_{mn} and α_{0n}.** In order to find α_{mn}, α_{0n} from the conditions (8.7e), (8.7c) we must make use of the values of $\overset{1}{\underset{4}{c}}_m$ found in 3. The result is up to the desired order

$$\alpha_{mn} = \{2\dot{\eta}^m\dot{\eta}^n + \delta_{mn}\tilde{\chi}\} \tag{14.1}$$

and

$$\alpha_{0n} = -\dot{\eta}^s\dot{\eta}^s\dot{\eta}^n + \tilde{\chi}\dot{\eta}^n - \tilde{\chi}\dot{\zeta}^n. \tag{14.2}$$

Finally from our last remark in 13 follows:

$$\alpha = 2\dot{\eta}^s\dot{\eta}^s + \tfrac{1}{2}\tilde{\chi}; \qquad \beta = 2\dot{\zeta}^s\dot{\zeta}^s + \tfrac{1}{2}\tilde{\psi}. \tag{14.3}$$

15. **Calculation of $\underset{6}{\Lambda}_{mn}$.** In the calculation of $\underset{6}{\Lambda}_{mn}$ for our present purposes, we may assume that $\underset{4}{c}_m$ is zero, as we shall now show.

After we have evaluated the surface integrals $\underset{6}{c}_m$, we may write the approximate equations of motion in the form

$$\lambda^4\underset{4}{c}_m + \lambda^6\underset{6}{c}_m = 0. \tag{15.1}$$

But this shows that when the motion is in accordance with (15.1) the quantities $\lambda^4\underset{4}{c}_m$ and $\lambda^6\underset{6}{c}_m$ will be of the same order of magnitude. It is evident, however, that $\lambda^4\underset{4}{c}_m$ can enter $\lambda^6\underset{6}{\Lambda}_{mn}$ only in combination with a quantity of the type $\lambda^2\underset{2}{\Theta}$. It will therefore enter only in terms which actually belong to the order λ^8 or higher, and since we do not propose to go beyond the order λ^6 in the calculation of the equations of motion, we may neglect all terms in $\underset{6}{\Lambda}_{mn}$ in which $\underset{4}{c}_m$ appears. Even if we make use of this fact, however, the calculations are still quite tedious, and there are actually forty-one different types of term in the expansion of $\underset{6}{\Lambda}_{mn}$. We find:

$$
\begin{aligned}
2\underset{6}{\Lambda}_{mn} = {} & -\underset{5}{\gamma}_{0m,0n} - \underset{5}{\gamma}_{0n,0m} + \delta_{mn}\underset{4}{\gamma}_{00,00} + \underset{4}{\gamma}_{mn,00} - \varphi\underset{4}{\gamma}_{00,mn} - \varphi\underset{4}{\gamma}_{ss,mn} - \varphi_{,mn}\underset{4}{\gamma}_{00} \\
& - \varphi_{,mn}\underset{4}{\gamma}_{ss} + \varphi_{,ms}\underset{4}{\gamma}_{ns} + \varphi_{,ns}\underset{4}{\gamma}_{ms} - \delta_{mn}\varphi_{,sr}\underset{4}{\gamma}_{sr} - 2\varphi_{,s}\underset{4}{\gamma}_{mn,s} + \varphi_{,s}\underset{4}{\gamma}_{ms,n} \\
& + \varphi_{,s}\underset{4}{\gamma}_{ns,m} - \tfrac{1}{2}\varphi_{,m}\underset{4}{\gamma}_{ss,n} - \tfrac{1}{2}\varphi_{,n}\underset{4}{\gamma}_{ss,m} - \tfrac{1}{2}\varphi_{,n}\underset{4}{\gamma}_{00,m} - \tfrac{1}{2}\varphi_{,m}\underset{4}{\gamma}_{00,n} \\
& + \tfrac{3}{2}\delta_{mn}\varphi_{,s}\underset{4}{\gamma}_{rr,s} + \tfrac{3}{2}\delta_{mn}\varphi_{,s}\underset{4}{\gamma}_{00,s} - \underset{3}{\gamma}_{0s}\underset{3}{\gamma}_{0n,ms} - \underset{3}{\gamma}_{0s}\underset{3}{\gamma}_{0m,ns} + 2\underset{3}{\gamma}_{0s}\underset{3}{\gamma}_{0s,mn} \\
& + \tfrac{1}{2}\delta_{mn}\underset{3}{\gamma}_{0s,r}\underset{3}{\gamma}_{0r,s} - \tfrac{3}{2}\delta_{mn}\underset{3}{\gamma}_{0s,r}\underset{3}{\gamma}_{0s,r} + \underset{3}{\gamma}_{0s,m}\underset{3}{\gamma}_{0s,n} + \underset{3}{\gamma}_{0m,s}\underset{3}{\gamma}_{0ns} - \varphi_{,0n}\underset{3}{\gamma}_{0m} \\
& - \varphi_{,0m}\underset{3}{\gamma}_{0n} + 2\delta_{mn}\varphi_{,0s}\underset{3}{\gamma}_{0s} - \varphi_{,0}\underset{3}{\gamma}_{0m,n} - \varphi_{,0}\underset{3}{\gamma}_{0n,m} - \varphi_{,n}\underset{3}{\gamma}_{0m,0} - \varphi_{,m}\underset{3}{\gamma}_{0n,0} \\
& + 2\varphi\underset{3}{\gamma}_{0m,0n} + 2\varphi\underset{3}{\gamma}_{0n,0m} - 2\delta_{mn}\varphi\varphi_{,00} + 2\varphi\varphi\varphi_{,mn} - \varphi\varphi_{,m}\varphi_{,n} \\
& + \tfrac{3}{2}\delta_{mn}\varphi\varphi_{,s}\varphi_{,s} + \tfrac{1}{2}\delta_{mn}\varphi_{,0}\varphi_{,0}\,.
\end{aligned}
\tag{15.2}
$$

The condition that $\underset{6}{\Lambda}_{mn,n}$ must be zero affords a valuable test of the correctness of the above formula. We have worked out the divergence of the $\underset{6}{\Lambda}_{mn}$ given in (15.2) and have found that it does indeed vanish.

16. **The Surface integrals for $l = 3$.** In order to find the principal deviation from the Newtonian laws of motion, all that essentially remains is to calculate the values of the surface integrals $\underset{6}{c}_m$. To do this we must first insert in (15.2) the values previously found for $\underset{4}{\gamma}_{00}$, $\underset{4}{\gamma}_{mn}$ and $\underset{5}{\gamma}_{0n}$ and then it is a matter of calculating the contributions of the resulting terms one by one and adding the expressions obtained. The general technique is similar to that used in 11 for the evaluation of $\underset{4}{c}_m$ but considerably more complicated.

On making use of our right to take $\underset{4}{c}_m$ to be zero, we may express the result in the form

$$
\begin{aligned}
\overset{1}{\underset{6}{c}}_m &= \frac{1}{4\pi}\int^{1} 2\underset{6}{\Lambda}_{mn}\cos(\mathbf{n}\cdot\mathbf{N})\,dS \\
&= -4\overset{1}{m}\overset{2}{m}\left\{\left[\dot\eta^s\dot\eta^s + \tfrac{3}{2}\dot\zeta^s\dot\zeta^s - 4\dot\eta^s\dot\zeta^s - 4\frac{\overset{2}{m}}{r} - 5\frac{\overset{1}{m}}{r}\right]\frac{\partial}{\partial\eta^m}\left(\frac{1}{r}\right)\right. \\
&\qquad \left.[4\dot\eta^s(\dot\zeta^m - \dot\eta^m) + 3\dot\eta^m\dot\zeta^s - 4\dot\zeta^s\dot\zeta^m]\frac{\partial}{\partial\eta^s}\left(\frac{1}{r}\right) + \frac{1}{2}\,\frac{\partial^3 r}{\partial\eta^s\,\partial\eta^r\,\partial\eta^m}\dot\zeta^s\dot\zeta^r\right\}.
\end{aligned}
\tag{16.1}
$$

17. **The Main Deviation from the Newtonian Equations of Motion.** In order to obtain the equations of motion belonging to this stage of our approximation, we must write

$$\lambda^4 \overset{k}{\underset{4}{c}}{}_m + \lambda^6 \overset{k}{\underset{6}{c}}{}_m = 0 \qquad k = 1,2 \tag{17.1}$$

and then must reabsorb the λ's by going over to the old time x^0 instead of the auxiliary time $\tau = \lambda x^0$ and by introducing a corresponding change in mass from m to M, where $M = \lambda^2 m$. There will be no confusion if we keep the old notation for the new quantities so that now $\dot{\xi} = d\xi/dx^0$ instead of $d\xi/d\tau$, and m is written for the new mass M. And with this convention we may write the equations of motion (17.1), by means of (11.10) and (16.1), in the form

$$\begin{aligned} \ddot{\eta}^m - \overset{2}{m}\frac{\partial(1/r)}{\partial\eta^m} = \overset{2}{m}\Bigg\{&\left[\dot{\eta}^s\dot{\eta}^s + \tfrac{3}{2}\dot{\zeta}^s\dot{\zeta}^s - 4\dot{\eta}^s\dot{\zeta}^s - 4\frac{\overset{2}{m}}{r} - 5\frac{\overset{1}{m}}{r}\right]\frac{\partial}{\partial\eta^m}(1/r) \\ &[4\dot{\eta}^s(\dot{\zeta}^m - \dot{\eta}^m) + 3\dot{\eta}^m\dot{\zeta}^s - 4\dot{\zeta}^s\dot{\zeta}^m]\frac{\partial}{\partial\eta^s}(1/r) + \frac{1}{2}\frac{\partial^3 r}{\partial\eta^s\,\partial\eta^r\,\partial\eta^m}\dot{\zeta}^s\dot{\zeta}^r\Bigg\}. \end{aligned} \tag{17.2}$$

The equations of motion for the other particle are obtained by replacing $\overset{1}{m}$, $\overset{2}{m}$, η, ζ by $\overset{2}{m}$, $\overset{1}{m}$, ζ, η.

These equations, giving the relativistic motion of two massive gravitating bodies, constitute the main result of our calculations from the point of view of practical application.

These equations have since been integrated by H. P. Robertson, whose results are given in the following note on "The Two Body Problem in General Relativity," Math. Ann. 39, p. 101 (1938).

We should like to thank Professor Robertson for the very kind interest he took in this problem and for his help.

THE INSTITUTE FOR ADVANCED STUDY.

Annals of Mathematics
Vol. 39, No. 1, January, 1938

NOTE ON THE PRECEDING PAPER: THE TWO BODY PROBLEM IN GENERAL RELATIVITY

By H. P. Robertson

(Received October 30, 1937)

In the preceding paper Einstein, Infeld and Hoffmann[1] have developed a most ingenious and useful method for obtaining, by successive approximations, the gravitational field and equations of motion of n bodies in the general theory of relativity. They have carried out the derivation to that order which leads, in the well known solution of the one body problem, to the perihelion advance of an infinitesimal planet, and have given explicitly the equations of motion for the case of two bodies of comparable masses.[2] It is the purpose of this note to integrate these latter equations, to the same approximation, emphasizing the effects on the orbit of a double star of possible astronomical interest.

Their equations of motion (17.2) differ from the classical equations of the two body problem by the appearance on the right of the specifically relativistic terms, of order m/r compared with those on the left. It is therefore expedient to introduce in place of the coordinates η_n, ζ_n the six variables

$$\alpha_n = \left(\overset{1}{m}\eta_n + \overset{2}{m}\zeta_n\right)/M, \qquad \xi_n = \eta_n - \zeta_n, \tag{1}$$

where M is the sum of the two masses m; we follow the classical terminology in referring to these as the coordinates of the center of gravity and of the relative orbit, respectively. Now to the approximation to which the equations (17.2) are valid, it suffices to replace the η_n, ξ_n in all terms on the right by their expressions in terms of the classical approximation $\alpha_n = a_n$, $\xi_n = x_n$ on which the relativistic solution is to be based. We take without further ado as this classical solution the elliptic motion defined by

$$\begin{gathered} a_n = 0, \qquad x_1 = r\cos w, \qquad x_2 = r \sin w, \qquad x_3 = 0, \\ u = r^{-1} = p^{-1}[1 + \epsilon \cos (w - \omega)]; \end{gathered} \tag{2}$$

the angular momentum and energy integrals are then given by

$$L \equiv r^2 \dot{w} = (Mp)^{\frac{1}{2}}, \qquad E \equiv \frac{v^2}{2M} - \frac{1}{r} = -\frac{1}{2a}, \tag{3}$$

and enable us to express the squares of the relative velocity v and its radial component $\dot r$ as simple polynomials in u.

---

[1] Ann. Math. **39**, p. 65 (1938). I am indebted to these authors for the opportunity of seeing their paper in manuscript, and for stimulating discussions.

[2] Op. cit., eqs. (17.2).

We are now in a position to determine the equations of motion for α_n, ξ_n, in a form suitable for integration, from the appropriate linear combinations of the equations (17.2) for η_n, ζ_n. Replacing

$$\eta_n \text{ by } \overset{2}{m}x_n/M, \qquad \zeta_n \text{ by } -\overset{1}{m}x_n/M,$$

in all terms on the right, and expressing v^2, $\dot{r}^2$ in terms of u with the aid of the integrals (3), we are led to the equations

$$\ddot{\alpha}_n = Ax_n + B\dot{x}_n, \qquad \ddot{\xi}_n + M\xi_n/\rho^3 = Cx_n + D\dot{x}_n, \tag{4}$$

in which $\rho^2 = \Sigma\xi^2$ and the coefficients are the scalar functions

$$\begin{aligned} A &= \frac{\overset{1}{m}\overset{2}{m}\delta m}{2Ma}u^3(1 - 4au + 3apu^2), \\ B &= -\frac{\overset{1}{m}\overset{2}{m}\delta m}{M^2}\dot{u}, \qquad \text{where } \delta m \equiv \overset{2}{m} - \overset{1}{m}, \\ C &= \frac{M^2}{2a}u^3([2, 7] + [4, 6]au - [0, 3]apu^2), \\ D &= -M[4, 6]\dot{u}, \qquad \text{where } M^2[k, l] \equiv k\overset{1}{m}{}^2 + l\overset{1}{m}\overset{2}{m} + k\overset{2}{m}{}^2. \end{aligned} \tag{5}$$

It is clear from (4) that the relative orbit is again plane in the ξ-space—for the x_n, $\dot{x}_n$ on the right may be replaced by ξ_n, $\dot{\xi}_n$—and we may therefore take

$$\alpha_3 = 0, \qquad \xi_3 = 0 \tag{6}$$

by a suitable small change of coordinates, without affecting the classical approximation (2). The quantities

$$\Lambda \equiv \xi_1\dot{\xi}_2 - \xi_2\dot{\xi}_1, \qquad \mathrm{E} \equiv \frac{1}{2M}(\dot{\xi}_1^2 + \dot{\xi}_2^2) - \frac{1}{\rho}, \tag{7}$$

then satisfy the equations

$$\frac{d\Lambda}{dt} = D(x_1\dot{x}_2 - x_2\dot{x}_1), \qquad \frac{d\mathrm{E}}{dt} = Cr\dot{r} + Dv^2; \tag{8}$$

with the aid of (3), (5) these equations may immediately be integrated, yielding

$$\begin{aligned} \Lambda &= L(1 - Mu[4, 6]), \\ \mathrm{E} &= E\{1 - Mu([6, 5] - [10, 15]au + [0, 1]apu^2). \end{aligned} \tag{9}$$

The differential equation of the orbit, in terms of the polar coordinates ρ, θ in the ξ_1, ξ_2 plane, is found by eliminating dt between the two integrals (9); on differentiating the resulting first order equation we obtain the more convenient linear differential equation

$$\frac{d^2v}{d\theta^2} + v = \frac{1}{p}\left\{1 - \frac{M}{2a}([2, 7] - [12, 18]au - [0, 3]apu^2)\right\}, \tag{10}$$

where $v = 1/\rho$. The polar equation of the relative orbit is then readily found by integrating (10) after replacing u by its expression (2) in terms of the angle $w = \theta$, and the time t at which the first mass is at a definite position θ in the

orbit can subsequently be obtained from the angular momentum integral by a quadrature.

The integrated form of the equations of motion could thus be obtained precisely to the order in question; however, the periodic terms due to the relativistic correction would scarcely be of astronomical interest, as they are still of the relative order m/p ($\sim 10^{-8}$ for a double star with mass comparable with that of the Sun, and a separation of one astronomical unit). We therefore determine explicitly only the *secular* perturbations of the velocity $\dot{\alpha}_n$ of the center of gravity and of the elements p, a, ω of the orbit—i.e., only the cumulative terms, which increase linearly with the time. For the velocity $\dot{\alpha}_n$ of the center of gravity this term is obtained from the temporal mean

$$\frac{1}{T}\int_0^T \ddot{\alpha}_n\, dt = \frac{1}{2\pi a^{\frac{3}{2}} p^{\frac{1}{2}}} \int_0^{2\pi} \ddot{\alpha}_n r^2\, dw, \tag{11}$$

over a period of the classical motion, of the acceleration $\ddot{\alpha}_n$. It is seen by inspection of the expressions (4), (5) that the only component of the acceleration which could contribute to a secular change in the velocity is that in the direction of the line of apsides; it further follows by direct integration of (11) that the various terms resulting from this component *exactly cancel in the mean.* A simple analysis of the "distance" of this center of gravity from a distant observer shows that this coordinate acceleration $\ddot{\alpha}_n$ is the same, to within periodic terms, as that which the observer would infer by the usual astronomical methods, and we may therefore conclude that the velocity of the center of gravity of a double star should show no secular change, of the order in question, due to this relativistic effect.

This result is to be contrasted with the surprising one obtained by Levi-Civita,[3] who finds a residual term

$$-\frac{\overset{1}{m}\overset{2}{m}\,\delta m\, \epsilon}{2Ma^{\frac{1}{2}}p^{\frac{5}{2}}}$$

in the acceleration in the direction of the periastron of the principal component of the double star. The discrepancy between the two results is presumably to be traced to the derivation of the equations of motion; a change in the numerical coefficients of the various terms on the right (in particular, of the interaction term referred to in the footnote above) will in general lead to an average acceleration of the above form. Such a result would be of considerable astronomical interest, as it would cause the double star as a whole to describe an

[3] Am. Journ. Math. **59**, p. 225 (1937). Also with the similar result implied by the equations of motion obtained by de Sitter (M. N. **77**, p. 162 [1916]; on taking these equations in the more readily comparable form given by J. Chazy (*La théorie de la relativité et la mécanique céleste*, t. 2, p. 177 [Gauthier-Villars, Paris, 1930]), they are found to lead to a mean acceleration five times as large as that obtained by Levi-Civita, and in the opposite direction. It may be of interest to note that Chazy's equations differ from (17.2) only in the absence of the fifth term on the right—an interaction term depending on the product of the two masses which is, as I am informed by Dr. Infeld, the term most difficult to determine.

approximately circular orbit, whose radius would be of the order a^2/m—which might amount to as much as a hundred parsecs!

We turn next to the relative orbit. It is seen immediately from (3), (9) that the parameter p and the semi-major axis a suffer no secular change; the shape of the orbit suffers no progressive change to this order. These results are also evident from the fact that the equations of motion to this order are indifferent to a change in sign of the time t; in particular, a secular change in a would imply a continuous loss of energy—compare the remarks at the beginning of the preceding paper concerning the radiation of energy in the form of gravitational waves.

Finally, we examine the equation of the orbit (10) with reference to the secular change in the longitude ω of periastron. Now the only relativistic term on the right of (10) which could lead to such an effect is the resonance term involving $\cos(\theta - \omega)$, for the others are either constants or contain $\cos 2(\theta - \omega)$, and can only lead to perturbations of the order M/r. The equation of the orbit may therefore be taken as

$$\frac{d^2 v}{d\theta^2} + v = \frac{1}{p}\left[1 + \frac{6M\epsilon}{p}\cos(\theta - \omega)\right], \tag{12}$$

on collecting the resonance terms arising from u and u^2. The integral of this equation may be written, to terms of the same order, in the form

$$v = \rho^{-1} = p^{-1}[1 + \epsilon\cos(\theta - \omega - \delta\omega)], \tag{13}$$

$$\text{where} \quad \delta\omega = \frac{3M}{p}\theta\left(= \frac{3GM}{c^2 p}\theta \text{ in c. g. s. units}\right).$$

It is to be noted that this is precisely the perihelion advance, predicted in the one body problem, of an infinitesimal planet in the field of a star of mass M.[4]

In summary, to this order the orbit of a double star in general relativity differs, in its secular behavior, from the classical orbit only in an advance of periastron equal to that which an infinitesimal planet, describing the same relative orbit, would undergo in the field of a star whose mass is the sum of those of the two components of the double star. The relative orbit is "plane," and its shape suffers only periodic perturbations of the negligible order M/r. It is to be noted, however, that we are here dealing with ideal spherically symmetric bodies, and that these results can accordingly be compared with observation only on allowing for possible oblateness or tidal effects of the one component on the other.

PRINCETON UNIVERSITY.

[4] In agreement with the result obtained by Levi-Civita, *op. cit.*, p. 230. The corresponding result on the de Sitter treatment, as computed from Chazy's equations, is the above multiplied by the factor $1 + [0, 5/3]$—and is, incidentally, the same as that announced by Levi-Civita in a preliminary report on his work at the Harvard Tercentenary Conference in September, 1936.

ANNALS OF MATHEMATICS
Vol. 39, No. 1, January, 1938

FINITE APPROXIMATIONS TO LIE GROUPS

BY A. M. TURING

(Received April 28, 1937; revised September 29, 1937)

A certain sense in which a finite group may be said to approximate the structure of a metrical group will be discussed. On account of Jordan's theorem on finite groups of linear transformations[1] it is clear that we cannot hope to approximate a general Lie group with finite subgroups. I shall show that we cannot approximate even with groups which are 'approximately subgroups': in fact the only approximable Lie groups are the compact Abelian groups. The key to the situation is again afforded by Jordan's theorem, but it is not immediately applicable. It is necessary to find representations of the approximating groups whose degree depends only on the group approximated.

Approximability of metrical groups. Suppose G is a group with a metric D invariant under left transformations i.e. $D(ax, ay) = D(x, y)$ for all x, y, a of G. Let H_ϵ be a finite subset of G in which is defined a second product with respect to which it forms a group (if a and b are in H_ϵ their product as elements of G will be written ab; the product of them as elements of H_ϵ will be written $a_0 b$, the inverse of a as an element of H_ϵ is written $[a]^{-1}$ and the identities of G and H_ϵ are written e, e_ϵ), and suppose each element x of G is within distance ϵ of an element $r(x)$ of H_ϵ, and for each a, b of H_ϵ $D(a_0 b, ab) > \epsilon$. Then H_ϵ will be said to be an ϵ-approximation to G.

A group is said to be approximable if it has an ϵ-approximation for each $\epsilon > 0$.

Immediately from the definition we see that an approximable group is totally bounded i.e. conditionally compact. It is therefore possible to find a metric which is both left and right invariant and equivalent to the given metric, in the sense that the class of open sets is the same for either metric. In future therefore we shall suppose that our metric is both ways invariant, and we shall denote the distance between x and y by $D(x, y)$.

It has been shown by J. v. Neumann[2] that with a conditionally compact

---

[1] This theorem states that a finite group of linear transformations has an Abelian self conjugate subgroup whose index does not exceed a certain bound depending only on the degree.

[2] J. v. Neumann, *Zum Haarschen Mass in topologischen Gruppen*, Compositio Mathematica, vol. 1 (1934) pp. 106–114; or alternatively, J. v. Neumann, *Almost periodic functions in a group*, Transactions of the American Mathematical Society, vol. 36 (1934) pp. 445–492, (remember that every continuous function in a conditionally compact group is a.p.). If the reader prefers to restrict the group in some way and use some other mean he has only to verify the inequality (1).

group we can define a mean for each continuous (complex-valued) function in the group, in such a way that (denoting the mean of $f(x)$ by $\int_G f(x)\,dx$)

$$\int_G (f(x) + g(x))\,dx = \int_G f(x)\,dx + \int_G g(x)\,dx$$

$$\int_G f(ax)\,dx = \int_G f(xa)\,dx = \int_G f(x)\,dx$$

and so that if $\epsilon > 0$ (and $f(x)$ is continuous), then there is a finite set of elements $a_1, a_2, \cdots, a_N$ of G such that

$$\left| \frac{1}{N} \sum_{i=1}^{N} f(xa_i) - \int_G f(x)\,dx \right| < \epsilon. \tag{1}$$

Before proceeding to the proofs of our main theorems we shall establish some elementary inequalities following immediately from our definition. Suppose the function $r(x)$ belongs to an ϵ-approximation H_ϵ to G, then

$$\begin{aligned} D(r(x)_0 r(y), xy) \leqq D(r(x)_0 r(y), r(x)r(y)) + D(r(x)r(y), xr(y)) \\ + D(xr(y), xy) < 3\epsilon \end{aligned} \tag{2}$$

and for any a, c of H_ϵ

$$D(c_0 a_0 [c]^{-1}, cac^{-1}) < 4\epsilon \tag{3}$$

for

$$\begin{aligned} D(c_0 a_0 [c]^{-1}, cac^{-1}) &< D(ca[c]^{-1}, cac^{-1}) + 2\epsilon \\ &= D([c]^{-1}c, e) + 2\epsilon \\ &\leqq D([c]^{-1}c, [c]_0^{-1}c) + D(e_\epsilon, e) + 2\epsilon \\ &\leqq D(e_\epsilon, e) + 3\epsilon \\ &= D(e_\epsilon^2, e_\epsilon) + 3\epsilon \\ &\leqq D(e_{\epsilon 0} e_\epsilon, e_\epsilon) + 4\epsilon = 4\epsilon. \end{aligned}$$

THEOREM 1. *Let G be an approximable group with a true continuous representation by matrices of degree n. Then it may be approximated by finite groups with true representations of the same degree n.*

LEMMA. *If H_η is an η-approximation (of order h_η) to the group G and if $f(x)$ is a continuous function in G such that*

$$|f(x) - f(x')| < \Delta \quad \text{when} \quad D(x, x') < \eta$$

then

$$\left| \frac{1}{h_\eta} \sum_{a \in H_\eta} f(a) - \int_G f(x)\,dx \right| \leqq 2\Delta. \tag{4}$$

We put

$$\int_G f(x)\,dx = A \qquad \frac{1}{h_\eta}\sum_{a \in H_\eta} f(a) = B$$

then given $\epsilon > 0$ there are $a_1, a_2, \cdots, a_N$ such that

$$\left| \frac{1}{N}\sum_{i=1}^{N} f(xa_i) - A \right| < \epsilon \tag{5}$$

for each element x of G. If in (5) we successively put x equal to each member of H_η and combine the resulting inequalities we obtain

$$\left| \frac{1}{Nh_\eta}\sum_{i=1}^{N}\sum_{c \in H_\eta} f(ca_i) - A \right| < \epsilon,$$

but $D(ca_i, c_0 r(a_i)) < 2\eta$, so that $|f(ca_i) - f(c_0 r(a_i))| < 2\Delta$ and therefore

$$\left| \frac{1}{Nh_\eta}\sum_{i=1}^{N}\sum_{c \in H_\eta} f(c_0 r(a_i)) - A \right| < \epsilon + 2\Delta. \tag{6}$$

However

$$\frac{1}{h_\eta}\sum_{c \in H_\eta} f(c_0 r(a_i)) = B$$

so that (6) yields (4) since ϵ was arbitrary.

PROOF OF THE THEOREM. Without loss of generality we may suppose that the given representation of G does not contain any irreducible component more than once. Let $\chi(x)$ be the character of the representation. This function will satisfy

$$\chi(x) = \int_G \chi(xy)\overline{\chi(y)}\,dy \tag{7}$$

$$\chi(x) = \chi(cxc^{-1}) \tag{8}$$

$$|\chi(x)| \leqq n \tag{9}$$

and since it is the character of a true representation

$$\chi(x) \neq \chi(e) = n \quad \text{if} \quad x \neq e.$$

Let $\epsilon > 0$. Then for some α, $1 > \alpha > 0$, $|\chi(x) - n| > \alpha$ when $D(x, e) \geqq \frac{1}{4}\epsilon$. Now let η be so chosen that $\epsilon/16 > \eta > 0$ and

$$|\chi(x) - \chi(x')| < \alpha/(50n^2) \quad \text{when} \quad D(x, x') < 4\eta \tag{10}$$

$$|\chi(ay)\overline{\chi(y)} - \chi(ay')\overline{\chi(y')}| < \alpha/(50n) \text{ all } a, \text{ when } D(y, y') < 2\eta \tag{11}$$

and take a corresponding η-approximation H_η. If we put

$$\varphi(a) = \frac{1}{h_\eta}\sum_{c \in H_\eta} \chi(c_0 a_0 [c]^{-1}) \tag{12}$$

then

$$(13) \qquad |\varphi(a) - \chi(a)| \leqq \frac{1}{h_\eta} \sum_{c \in H_\eta} |\chi(c_0 a_0 [c]^{-1}) - \chi(cac^{-1})| < \frac{\alpha}{50n^2}$$

for $D(c_0 a_0 [c]^{-1}, cac^{-1}) < 4\eta$ by (3) and therefore each summand is less than $\alpha/(50n^2)$. We have

$$(14) \qquad \begin{aligned} \left| \frac{1}{h_\eta} \sum_{b \in H_\eta} \varphi(a_0 b)\overline{\varphi(b)} - \chi(a) \right| &\leqq \frac{1}{h_\eta} \left| \sum_{b \in H_\eta} (\varphi(a_0 b)\overline{\varphi(b)} - \chi(a_0 b)\overline{\chi(b)}) \right| \\ &+ \frac{1}{h_\eta} \left| \sum_{b \in H_\eta} (\chi(a_0 b) - \chi(ab))\overline{\chi(b)} \right| + \left| \frac{1}{h_\eta} \sum_{b \in H_\eta} \chi(ab)\overline{\chi(b)} - \int_G \chi(ay)\overline{\chi(y)}\, dy \right|. \end{aligned}$$

Applying the lemma to $\chi(ay)\overline{\chi(y)}$ and making use of (11) we have

$$(15) \qquad \left| \frac{1}{h_\eta} \sum_{b \in H_\eta} \chi(ab)\overline{\chi(b)} - \int_G \chi(ay)\overline{\chi(y)}\, dy \right| < \frac{2\alpha}{50n}$$

and from (9), (10) we obtain

$$(16) \qquad \frac{1}{h_\eta} \left| \sum_{b \in H_\eta} (\chi(a_0 b) - \chi(ab))\overline{\chi(b)} \right| < \frac{\alpha}{50n}.$$

Finally

$$(17) \qquad \begin{aligned} &\frac{1}{h_\eta} \left| \sum_{b \in H_\eta} (\varphi(a_0 b)\overline{\varphi(b)} - \chi(a_0 b)\overline{\chi(b)}) \right| \\ &\leqq \frac{1}{h_\eta} \sum_{b \in H_\eta} |(\varphi(a_0 b) - \chi(a_0 b))\overline{\varphi(b)}| + \frac{1}{h_\eta} \sum_{b \in H_\eta} |(\overline{\varphi(b)} - \overline{\chi(b)})\chi(a_0 b)| \\ &< \frac{2\alpha}{50n} \end{aligned}$$

by (9) and (13). Combining (14), (15), (16), (17),

$$(18) \qquad \left| \frac{1}{h_\eta} \sum \varphi(a_0 b)\overline{\varphi(b)} - \chi(a) \right| < \frac{\alpha}{10n}$$

$$(19) \qquad \left| \frac{1}{h_\eta} \sum \varphi(a_0 b)\overline{\varphi(b)} - \varphi(a) \right| < \frac{\alpha}{8n}.$$

Now $\varphi(a) = \varphi(c_0 a_0 [c]^{-1})$ for each a, c of H_η. This function is therefore expressible as a sum of characters

$$\varphi(a) = \sum_{\lambda=1}^{M} \alpha_\lambda \chi^{(\lambda)}(a),$$

$\chi^{(1)}(a), \cdots, \chi^{(M)}(a)$ being the characters of the different irreducible representations of H_η. From the general theory of representations

$$\frac{1}{h_\eta} \sum_{b \in H_\eta} \chi^{(\lambda)}(a_0 b)\overline{\chi^{(\mu)}(b)} = \delta_{\lambda\mu} \chi^{(\lambda)}(a)$$

(19) therefore becomes

$$\left| \sum_{\lambda=1}^{M} \alpha_\lambda(\bar{\alpha}_\lambda - 1)\chi^{(\lambda)}(a) \right| < \frac{\alpha}{8n}.$$

Squaring each side of this inequality and summing over H_η,

$$\frac{1}{h_\eta}\sum_{\lambda=1}^{M}\sum_{a \in H_\eta} |\alpha_\lambda|^2 |1-\alpha_\lambda|^2 |\chi^{(\lambda)}(a)|^2 = \sum_{\lambda=1}^{M} |\alpha_\lambda|^2 |1-\alpha_\lambda|^2 < \frac{\alpha^2}{64n^2}.$$

If we define $\xi(a)$ by

$$\xi(a) = \sum_{|1-\alpha_\lambda|>|\alpha_\lambda|} \chi^{(\lambda)}(a)$$

it will satisfy

$$\frac{1}{h_\eta}\sum_{a \in H_\eta} \xi(a_0 b)\overline{\xi(b)} = \xi(a) \tag{20}$$

and

$$\begin{aligned} \frac{1}{h_\eta}\sum_{a \in H_\eta} |\xi(a) - \varphi(a)|^2 &= \sum_{\lambda=1}^{M} \operatorname{Min}\left(|\alpha_\lambda|^2, |1-\alpha_\lambda|^2\right) \\ &\leqq 4\sum_{\lambda=1}^{M} |\alpha_\lambda|^2 |1-\alpha_\lambda|^2 < \frac{\alpha^2}{16n^2}. \end{aligned} \tag{21}$$

We now wish to infer from the inequality (21) that the functions $\varphi(a)$ and $\xi(a)$ differ only slightly at each point of H_η. This is possible on account of the relations (19), (20).

$$\begin{aligned} &\left| \frac{1}{h_\eta}\sum_{b \in H_\eta} \left(\xi(a_0 b)\overline{\xi(b)} - \varphi(a_0 b)\overline{\varphi(b)}\right) \right| \\ &\leqq \sum_{b \in H_\eta} \frac{1}{h_\eta} \left| (\xi(a_0 b) - \varphi(a_0 b))\overline{\xi(b)} \right| + \sum_{b \in H_\eta} \frac{1}{h_\eta} \left| (\overline{\xi(b)} - \overline{\varphi(b)})\varphi(a_0 b) \right| \\ &\leqq \left\{\frac{1}{h_\eta}\sum_{b \in H_\eta} |\xi(b) - \varphi(b)|^2\right\}^{\frac{1}{2}} \left(\left\{\frac{1}{h_\eta}\sum_{b \in H_\eta} |\xi(b)|^2\right\}^{\frac{1}{2}} + \left\{\frac{1}{h_\eta}\sum_{b \in H_\eta} |\varphi(b)|^2\right\}^{\frac{1}{2}}\right) \\ &< \frac{\alpha}{4n}(n + n) = \frac{1}{2}\alpha, \end{aligned} \tag{22}$$

since $|\xi(b)| \leqq n$ and $|\varphi(b)| \leqq n$ for each b of H_η. Now combine (18), (20), (22), and we have

$$|\xi(a) - \chi(a)| < \frac{1}{2}\alpha + \frac{\alpha}{10n} < \alpha.$$

This implies that $\xi(e_\eta) = \chi(e) = n$ and that if $D(a, e) \geqq \frac{1}{4}\epsilon$ then $\xi(a) \neq \chi(e) = \xi(e_\eta)$. $\xi(a) = \xi(e_\eta)$ only for elements of a certain self-conjugate subgroup N entirely contained within distance $\frac{1}{4}\epsilon$ of the identity of G. The factor group has a true representation of degree n, and I shall show that it can be taken as a

ϵ-approximation to G. We choose an element in each coset of N as a representative of that coset and define the function $v(a)$ (a in H_η) to be the representative of the coset in which a lies. The totality of elements $v(a)$ we call K. Putting $v(a) \otimes v(b) = v(a_0 b)$, K forms a group with respect to the product $\otimes$. For each a of H_η there is an element m of N for which $v(a) = a_0 m$ and therefore

$$D(a, v(a)) \leqq D(a, am) + D(am, a_0 m) < \tfrac{1}{4}\epsilon + \eta.$$

Consequently if we put $R(x) = v(r(x))$ we have

$$D(R(x), x) \leqq D(v(r(x)), r(x)) + D(r(x), x) < (\tfrac{1}{4}\epsilon + \eta) + \eta < \epsilon$$

and

$$D(v(a) \otimes v(b), v(a)v(b)) \leqq D(v(a_0 b), a_0 b) + D(a_0 b, ab) + D(ab, v(a)v(b))$$
$$< 3(\tfrac{1}{4}\epsilon + \eta) + \eta < \epsilon,$$

which shows that K is an ϵ-approximation to G.

THEOREM 2. *An approximable Lie group is compact and Abelian.*

LEMMA. *A closed subgroup of a connected group cannot have a finite index greater than* 1.

Suppose H is a closed subgroup of G and has index i, $1 < i < \infty$. Then $G - H$ is not void and is closed, being the sum of a finite number of closed sets, the cosets of H. G is the sum of two closed disjoint sets neither of which is void, and therefore is not connected.

If G is a compact Lie group it cannot have a closed subgroup of positive measure different from the whole group.

PROOF OF THE THEOREM. An approximable Lie group is complete and conditionally compact, i.e. it is compact, and is therefore a group of linear transformations,[3] of degree n say. By theorem 1 we can approximate it by finite groups H_ϵ of linear transformations of degree n. But by Jordan's theorem[4] each finite group of linear transformations has an Abelian subgroup whose index does not exceed a certain bound $Z(n)$ depending only on the degree. Let A_ϵ be this Abelian subgroup in H_ϵ. Then there is a finite number $c_1, c_2, \cdots c_N$ ($N \leqq Z(n)$) of elements of H_ϵ such that every element of H_ϵ is of the form $c_{i0} a$ where a is in A_ϵ. For any x of G we have

$$D(x, r(x)) < \epsilon$$

$$r(x) = c_{i0} a,\ a \in A_\epsilon,\ i \leqq N$$

$$D(c_{i0} a, c_i a) < \epsilon.$$

---

[3] J. v. Neumann, *Die Einführung analytischer Parameter in topologischen Gruppen*, Annals of Mathematics, vol. 34 (1933), pp. 170–190.

[4] A. Speiser, *Theorie der Gruppen von endlicher Ordnung*, (Berlin 1927) 2nd ed., p. 215.

Hence every element of G is of the form $c_i a d$ where d is within distance 2ϵ of the identity of G and $i \leqq N$. The points ad must therefore form a set E_ϵ of measure $1/Z(n)$ at least. Now put $x = ad$ $y = a'd'$:

$$\begin{aligned} D(xy, yx) &= D(ada'd', a'd'ad) \\ &\leqq 2D(d, d') + D(a_0 a', a_0' a) + D(aa', a_0 a') + D(a_0' a, a'a) \qquad (23) \\ &< 6\epsilon. \end{aligned}$$

In the product group $G \times G$ we have therefore a set $E_\epsilon \times E_\epsilon$ of pairs (x, y) of measure $1/(Z(n))^2$ at least, in which $D(xy, yx) < 6\epsilon$. Now take a sequence ϵ_i tending to 0, and put $F_i = \sum_{j \geqq i} E_{\epsilon_j} \times E_{\epsilon_j}$, $E = \prod F_i$. For each $i \leqq N$,

$$mF_i \geqq m(E_{\epsilon_i} \times E_{\epsilon_i}) \geqq \frac{1}{(Z(n))^2}.$$

Then $mE \geqq 1/(Z(n))^2$ since the F_i are a decreasing sequence. If $(x, y) \epsilon E$ then for each i, $(x, y) \epsilon F_i$, i.e. $x \epsilon E_{\epsilon_j}$, $y \epsilon E_{\epsilon_j}$ for some $j \geqq i$. Then by (23), $D(xy, yx) < 6\epsilon_j \leqq 6\epsilon_i$: but i was arbitrary so that $D(xy, yx) = 0$, $xy = yx$.

Now let N_x be the set of those y for which $xy = yx$, i.e. the normaliser of x. Then

$$\int_G mN_x \, dx \geqq mE \geqq \frac{1}{(Z(n))^2}.$$

Consequently $mN_x > 0$ in an x-set of positive measure. But if $mN_x > 0$ we have $N_x = G$ by the lemma, for N_x is certainly closed. This shows that the centre of G is of positive measure, and again applying the lemma we see that G is Abelian.

PRINCETON UNIVERSITY.

Annals of Mathematics
Vol. 39, No. 1, January, 1938

LATTICES AND TOPOLOGICAL SPACES

By Henry Wallman

(Received July 31, 1937)

It is a classical result of topology that the homology theory of a geometrical polyhedron is determined by the abstract complex associated with any triangulation of the polyhedron. The abstract complex, of course, is abstractly simpler than the polyhedron. This is a first instance of the fact that although topological invariants are attached to spaces, the invariants may actually be determined by structures which are axiomatically simpler than the spaces themselves. Guided by this principle we prove in this paper that *the homology theory of very general spaces* (T_1-spaces in fact) *is determined by the distributive lattice of the closed sets of the space.*

Suppose a topological invariant I of a general space R is shown to be determined by a system L associated with R which is of simpler structure than R itself. Since the conditions on L are less stringent than those on R, there exist many spaces which have equivalent L's. Among these there may exist a space S which is of more special type than R. This would have the consequence that the invariant I is not selective enough to distinguish between the general spaces R and the more special spaces S. Thus we show here that *given any* (T_1)*-space there is a bicompact space which has the same homology theory*, and therefore that it is not possible by means of the homology theory to distinguish between general spaces and bicompact spaces.

The method used in this part is closely related to that of M. H. Stone in his recent work [5]. The differences arise from the approach to the subject. Stone's approach was the algebraic one which developed from the classification of the representations of Boolean rings, which necessitated a most detailed and exhaustive algebraic analysis. In this paper, however, the approach is strictly geometrical and this allows us to proceed very quickly to the homology theory of general spaces.

The results of Part I are then generalized in that we now start with a structure even simpler than a distributive lattice, namely an abstract multiplicative complex. Such an abstract complex is the nerve, in Alexandroff's sense, of a multiplicative infinite covering of a space (cf. [1], p. 257). We show how the complex K generates a bicompact space S. The relation between the space S and the complex K is that the nerve of a certain fundamental family of the closed sets of S is isomorphic to K. This fundamental family is what we call a sub-basis for the closed sets of S, that is a collection of sets whose finite sums form a multiplicative basis for the closed sets of S. An example of a closed sub-

basis in the plane is the set of all closed rectangles whose sides are parallel to the axes.

We are now able to define, entirely in terms of the simplicial structure of K, a notion of "covering" in K, such that these coverings in K correspond to the ordinary point-set coverings of S. We are also able to introduce a partial ordering of the vertices of K, which corresponds to the inclusion relation in S. It is quite evident that employing Čech's method ([3], Chap. II), a homology theory of the infinite complex can be defined by considering the finite coverings of K and by defining a covering $\mathfrak{V}$ in K to be a refinement of another covering $\mathfrak{U}$ if and only if each vertex of $\mathfrak{V}$ is less than, with respect to the partial ordering in K, some vertex of $\mathfrak{U}$. This theory we shall call the homology theory of K.

Now we consider the topological homology theory of the space S defined by the finite coverings by (arbitrary) closed sets ([3] p. 179). We show that if S is a Hausdorff space, this theory is identical with that determined by the finite coverings made up of closed sets of the sub-basis. Thus we have proved that *an abstract multiplicative complex generates a bicompact space, and the homologies of the complex and the space are identical.*

I. Lattices and Bicompact Spaces

1. Let L be a distributive lattice with zero element and unit element; that is, a collection of elements $a, b, c, \cdots$ together with operations of finite addition (indicated by the sign $+$) and finite multiplication (indicated by juxtaposition) such that L is closed under each of these operations, each operation is commutative:

$$a + b = b + a \qquad ab = ba,$$

associative:

$$a + (b + c) = (a + b) + c \qquad a(bc) = (ab)c,$$

idempotent:

$$a + a = a \qquad aa = a,$$

each operation distributes with respect to the other:

$$a(b + c) = ab + ac \qquad a + (bc) = (a + b)(a + c),$$

the statements $a + b = a$ and $ab = b$ are equivalent (either of the statements may be written $a > b$), and there exist elements 0 and 1 such that $a + 0 = a$ and $a1 = a$ for every a.

2. Definition. A collection $\mathfrak{S}$ of elements of L which has first the *finite intersection property*: "the product of any finite number of elements of the collection is not zero," and which is second a proper sub-collection of no collection having the finite intersection property, is called a *maximal* collection.

3. Lemma 1. *Given any collection $\mathfrak{F}$ of elements of L with the finite intersection property, there is a maximal collection $\mathfrak{G}$ containing $\mathfrak{F}$.*

Proof.[1] Suppose $\mathfrak{F} = \mathfrak{F}^1$ is not maximal. Then there is an $\mathfrak{F}^2$ with the finite intersection property properly containing $\mathfrak{F}^1$. Suppose $\mathfrak{F}^\alpha$ has been found for every ordinal $\alpha <$ some ordinal β so that $\mathfrak{F}^\alpha$ has the finite intersection property and properly contains $\mathfrak{F}^{\alpha'}$ for $\alpha' < \alpha$. If β has an immediate predecessor then if $\mathfrak{F}^{\beta-1}$ is maximal, let $\mathfrak{G} = \mathfrak{F}^{\beta-1}$; and if $\mathfrak{F}^{\beta-1}$ is not maximal, let $\mathfrak{F}^\beta$ be a collection with the finite intersection property properly containing $\mathfrak{F}^{\beta-1}$. If β is a limit ordinal, let $\mathfrak{F}^\beta$ be the collection sum of all $\mathfrak{F}^\alpha$, $\alpha < \beta$. Then $\mathfrak{F}^\beta$ has the finite intersection property and properly contains $\mathfrak{F}^\alpha$. But L has a definite cardinal number. Hence the process of obtaining a sequence of properly increasing collections with the finite intersection property must stop, and the last element of this sequence is $\mathfrak{G}$.

Remark. The particular nature of the finite intersection property is irrelevant to the proof. What is used is that this property imposes a restriction on *finite* subsets only of $\mathfrak{G}$.

4. Lemma 2. *A maximal collection is characterized by the properties:*

1) *If $a, b \in \mathfrak{G}$ then $ab \in \mathfrak{G}$.*

2) *$c \in \mathfrak{G}$ if and only if $ac \neq 0$ for every $a \in \mathfrak{G}$.*

Proof. Let $\mathfrak{G}$ be a maximal collection. If $g_1, \cdots, g_k$ are any finite number of elements of $\mathfrak{G}$, and if $a \in \mathfrak{G}$, $b \in \mathfrak{G}$, then $a, b, g_1, \cdots, g_k$ are also a finite number of elements of $\mathfrak{G}$ and hence $abg_1 \cdots g_k \neq 0$, so that $ab \in \mathfrak{G}$. $c \in \mathfrak{G}$ means that for every finite number $g_1, \cdots, g_n$ of elements of $\mathfrak{G}$, $cg_1 \cdots g_n \neq 0$. But by condition 1) already proved, $g_1 \cdots g_n$ is in $\mathfrak{G}$; call it a. Thus $\mathfrak{G}$ satisfies 1) and 2).

The converse is evident.

Corollary. *A maximal collection is a divisorless additive ideal*[2] *which does not contain zero, and conversely.*

Proof. That a maximal collection is an additive ideal which does not contain zero is obvious. To show that it is divisorless, suppose N is an additive ideal properly containing $\mathfrak{G}$, so that there is an element c which is in N but not in $\mathfrak{G}$. Then by condition 2) of Lemma 2 there is an $a \in \mathfrak{G}$ such that $ac = 0$. Hence N contains 0 and therefore, since it is additive and $x > 0$ for every $x \in L$, it coincides with L.

Conversely, let $\mathfrak{G}$ be a divisorless additive ideal not containing zero. Then $\mathfrak{G}$ evidently satisfies condition 1) of Lemma 2. If $c \in \mathfrak{G}$ then for every $a \in \mathfrak{G}$

---

[1] The proof is the one which is usual in demonstrating the existence of maximal collections of various types; cf. [4], p. 474, footnote 6.

[2] A subcollection M of L is an *additive ideal* if $a, b \in M$ imply $ab \in M$ and if $c \in M$ implies that $d \in M$ for every $d \in L$ such that $d > c$. An ideal which is a proper subideal of the entire lattice L only is called *divisorless*.

we must have $ac \neq 0$, else $\mathfrak{G}$ would contain 0. If $ac \neq 0$ for every $a \epsilon \mathfrak{G}$ then $c \epsilon \mathfrak{G}$, since otherwise the additive ideal containing c and $\mathfrak{G}$ would be an ideal properly containing $\mathfrak{G}$ which did not contain 0. Thus $\mathfrak{G}$ satisfies condition 2) of Lemma 2 also and hence is a maximal collection.

5. DEFINITION. A maximal collection of elements of L, or equivalently, a divisorless additive ideal not containing zero, is called a *point*, and the elements of L in the maximal collection are called the *coördinates* of the point.

6. DEFINITION. With each element of L we associate a set of points, called a *basic* set, consisting of all points which have the given element of L as one of their coördinates. The basic set associated with the element f of L is called the *basic f-set.*

REMARK 1. With the zero element of L is associated the zero set of points, and with the unit element of L is associated the unit set of points.

REMARK 2. A point belongs to the basic set determined by each of its coördinates, and since a point is a maximal collection, it is the only point common to the basic sets determined by all its coördinates, i.e., *a point is the intersection of the basic sets determined by its coördinates.*

LEMMA 3. *The necessary and sufficient condition that different elements of L determine different basic sets, i.e., that the correspondence between elements of L and basic sets be* 1:1, *is that L have the disjunction property*: "*if a and b are different elements of L there is an element c of L such that one of ac and bc is zero and the other is not zero.*"

PROOF. Suppose L has the disjunction property. Then (as we may assume as a matter of notation) $ac = 0$ and $bc \neq 0$. Hence, by Lemma 1, there is a point P one of whose coördinates is b and another of whose coördinates is c. P is in the basic b-set but not in the basic a-set (since $ac = 0$), proving that the basic a-set and the basic b-set are different.

Conversely, if the basic a-set is different from the basic b-set there is (again as a matter of notation) a point P in the basic b-set which is not in the basic a-set. Hence by condition 2) of Lemma 2 there is a coördinate c of P such that $ac = 0$; moreover, bc is not zero since any two coördinates of a point have a non-zero product. Thus L has the disjunction property.

7. LEMMA 4. *For any elements a, b of L the intersection of the basic a-set and the basic b-set is the basic ab-set.*

PROOF. Let P be a point. Since P is an additive ideal the two statements: both a and b are coördinates of P, ab is a coördinate of P, are equivalent.

COROLLARY. *If $ab = 0$, then the basic a-set and the basic b-set are disjoint.*

LEMMA 5. *For any elements a, b of L the sum of the basic a-set and the basic b-set is the basic $a + b$-set.*

PROOF. Let P be a point. Since P is a divisorless additive ideal it is also

a prime additive ideal.[3] Hence the two statements: either a or b is a coördinate of P, $a + b$ is a coördinate of P, are equivalent.

8. DEFINITION. The intersection (of a finite or infinite number) of basic sets is called a *closed* set.

LEMMA 6. *The closed sets satisfy the conditions:*

1) *The zero set and the unit set are closed sets.*

2) *The sum of two closed sets is a closed set.*

3) *The intersection (of a finite or infinite number) of closed sets is a closed set.*

4) *A point is a closed set.*

PROOF. 3) is obvious. 1) and 4) follow immediately from Remarks 1 and 2 of §6. To prove 2) suppose that A and B are closed sets, $A = \prod A_i$ and $B = \prod B_j$, where A_i and B_j are basic sets. Then $A + B = \prod (A_i + B_j)$, and since each $A_i + B_j$ is a basic set by Lemma 5, $A + B$ is itself a closed set.

9. Lemma 6 shows that the set of all points, with the definition of closed set of §8, forms a topological space, and in fact, a T_1-space.[4] Denote this space[5] by S.

LEMMA 7. *S is a bicompact T_1-space.*

PROOF. Given a collection $\mathfrak{A}$ of closed sets of S, any finite number of which have a non-zero intersection, we have to display a point common to all the closed sets of $\mathfrak{A}$. Replace each closed set of $\mathfrak{A}$ by the collection of basic sets of which it is the intersection, and call this collection of basic sets $\mathfrak{B}$. Then any finite number of basic sets of $\mathfrak{B}$ have a non-zero intersection. Let $\mathfrak{F}$ be the collection of all elements of L whose corresponding basic sets are in $\mathfrak{B}$. Then by the corollary to Lemma 4, the product of any finite number of elements of $\mathfrak{F}$ is not zero. Hence we may apply Lemma 1 to obtain a point common to all the basic sets of $\mathfrak{B}$ and thus common to all the closed sets of $\mathfrak{A}$.

10. Now let L be the distributive lattice with zero and unit of the closed sets of a T_1-space R under the operations of point set addition and intersection,[6]

---

[3] A *prime additive ideal* M is an additive ideal for which $c + d \in M$ implies that either $c \in M$ or $d \in M$. The proof that a divisorless additive ideal is also a prime additive ideal is this: Suppose that neither c nor d is in the divisorless additive ideal M. Then by Lemma 2 and its corollary there exist elements a and b in M such that $ac = 0$ and $bd = 0$. Hence $(c + d)ab = abc + abd = 0$, so that $c + d$ is not in M either.

[4] A space in which a point is closed, [2], p. 59.

[5] What we have done is to turn the set of points into a T_1-space by using the basic sets as a *multiplicative* basis for the *closed* sets of S. But we might have proceeded differently. For, the intersection of two basic sets is a basic set (Lemma 4); and if P and Q are distinct points there are disjoint basic sets containing them (condition 2) of Lemma 2). These facts would allow us to turn the set of points into a *Hausdorff* space T (not bicompact) by using the basic sets as an *additive* basis for the *open* sets of T. But from condition 2) of Lemma 2 it also follows that if P is a point not in the basic set F, there is a basic set G containing P such that $FG = 0$. Hence T would be totally-disconnected.

[6] L, in this case, has the disjunction property.

and let S be the bicompact T_1-space obtained from L by the procedure of the preceding paragraphs. It is easily seen that the collection of all closed sets of R containing a given point p of R (p itself is one such closed set) is a maximal collection and hence determines a certain point of S; such a point of S is called an *ordinary* point. Conversely let P be a point of S. If the closed sets of R which are the coördinates of P have a point p of R in common (they surely have not more than one point in common), then p itself is one of the coördinates of P, so that P consists of the set of *all* closed sets of R containing p; hence P is an ordinary point.[7] But if the closed sets of R which are the coördinates of P have no point of R in common, P is called an *ideal* point.

Remark 1. If R is bicompact, every point of S is ordinary. For the coördinates of a point of S have the property that any finite number have a non-zero intersection, and if R is bicompact there is a point p of R common to all these coördinates.

Remark 2. If p and q are different points of R then the ordinary points of S associated with them are different. For p and q are themselves coördinates of the points in S associated with them. Hence *the correspondence between the points of R and the ordinary points of S is* 1:1.

11. Denote the subspace of all ordinary points of S by R'.

Lemma 8. *R and R' are homeomorphic.*[8]

Proof. Denote the set of all ordinary points associated with a set h of R by the *ordinary h-set*. Let P be a point of R' and p the point of R with which it is associated. Let f be a closed set of R. Then if p is in the ordinary f-set, P is in the basic f-set since $f > p$ and P is an additive ideal. Conversely, if P is in the basic f-set, f is a coördinate of P and therefore $pf \neq 0$ and $p \epsilon f$, so that P is in the ordinary f-set. That is, *if f is a closed set of R then the ordinary f-set $= R' \cdot$ the basic f-set.* But this statement shows that the 1:1 correspondence between the points of R and R' establishes also a 1:1 correspondence between the closed sets of R and a basis for the closed sets of R', which proves the lemma.

12. The closure in S of a subset F of S is the intersection of all basic sets containing F, since any closed set containing F can be expressed as the intersection of basic sets containing F. Now suppose f is a closed set of R and F is the ordinary f-set. Then evidently the basic f-set contains F and if g is any closed set of R such that the basic g-set contains F we must have $g > f$. Hence the intersection of all basic sets containing F is the basic f-set. That is, *if f is a closed set in R then the closure in S of the ordinary f-set is the basic f-set.*

[7] Evidently the ordinary points are precisely those points of S which make up the basic sets determined by the individual points of R.

[8] From this Lemma and Remark 1 of §11 it follows that if R is bicompact S is homeomorphic to R, i.e., this process of bicompactification yields nothing new when applied to bicompact T_1-spaces. This is perhaps unfortunate since bicompact T_1-spaces which are not Hausdorff spaces need not be absolutely closed.

LEMMA 9. *R' is dense in S. If f and g are closed disjoint sets in R, and F and G their correspondents in S under the homeomorphism $R \rightleftarrows R'$, and $\bar{F}$ and $\bar{G}$ the closures in S of F and G, then $\bar{F}$ and $\bar{G}$ are also disjoint.*

PROOF. R' is the ordinary R-set. Hence its closure is the basic R-set, that is S.

If f, g are as in the hypothesis, $\bar{F}$ and $\bar{G}$ are the basic f-set and the basic g-set, and by the corollary to Lemma 4, these are disjoint.

13. It is convenient now to denote by the *basic open u-set, u* being an open set in R, the set of all points of S each of whose coördinates has a non-zero product with u.

Let f be a closed subset of R and let u be its complement in R, $u = R - f$. Let P be a point of S which is not in the basic f-set. Then by condition 2) of Lemma 2 there exists a coördinate a of P such that $af = 0$, so that a is contained in u. Since every coördinate of P has a non-zero product with a, every coördinate of P has a non-zero product with the larger set u, and consequently P is in the basic open u-set. Conversely, if P is in the basic open u-set, that is, if every coördinate of P has a non-zero product with u, then f is not a coördinate of P since $fu = 0$, which means that P is not in the basic f-set. Thus we have verified that *if f is a closed set in R and u its complement in R, the complement in S of the basic f-set is the basic open u-set.*

The basic f-sets form a (multiplicative) basis for the closed sets of S. Hence their complements, which by the paragraph above are the basic open sets, form an (additive) basis for the open sets of the same space S. From Lemmas 4 and 5 it follows that the sum of the basic open u-set and the basic open v-set is the basic open $u + v$-set, and the intersection of the basic open u-set and the basic open v-set is the basic open uv-set. Thus *the correspondence between open sets of R and basic open sets of S establishes an isomorphism (with respect to intersection relations) between the finite coverings by open sets of R and the finite coverings by basic open sets of S.* But in a bicompact space every finite covering by open sets has a refinement which is a finite covering by basic open sets. Hence[9]

LEMMA 10. *In the sense of Čech, the homology theory of R is identical with that of S.*

LEMMA 11. *In the sense of Čech,[10] dimension R = dimension S.*

PROOF. It is clear from the isomorphism of the finite coverings by open sets of R and the finite coverings by basic open sets of S that dimension $S \leqq$ dimension R. To show the converse, observe that there is a finite covering by open sets, $\mathfrak{n}$, of R none of whose refinements are of dimension less than dimension $R = n$. Let $\mathfrak{N}$ be the finite covering by basic open sets of S corre-

[9] Since the homology theory of a space is determined by a complete family of finite coverings by open sets. Cf. [3], p. 165.

[10] The dimension of a *covering* is the dimension of its nerve. The dimension of a *space* is the least integer n such that every finite covering by open sets has a refinement of dimension $\leqq n$.

sponding to $\mathfrak{n}$. Then no refinement of $\mathfrak{N}$ is of dimension $< n$. For if a refinement finite covering by open sets of S were of dimension $< n$, the set of ordinary points in its open sets would form a finite covering by open sets of R' of dimension $< n$, and this would give rise to a refinement of $\mathfrak{n}$ of dimension $< n$. Thus dimension S is not $<$ dimension R, so that dimension R = dimension S.

14. Lemma 12. *The necessary and sufficient condition*[11] *for S to be a Hausdorff space* (and hence normal) *is that R be normal.*[12]

Proof. Suppose R is normal. Let P and Q be distinct points of S. Then by condition 2) of Lemma 2 there exists a coördinate f of P and a coördinate g of Q such that $fg = 0$. Since R is normal there exist disjoint open sets u and v containing f and g respectively. Then the basic open u-set contains the basic u-set and hence the point P; the basic open v-set and the basic open v-set are disjoint. Therefore S is a Hausdorff space.

Conversely, suppose S is a Hausdorff space. Since S is bicompact it is normal.

In general, if A is any topological space imbedded in any normal space B in such a manner that if two closed sets in A are disjoint their closures in B are also disjoint, then A is itself normal. For let F and G be closed disjoint sets of A and $\bar{F}$ and $\bar{G}$ their closures in B. Since $\bar{F}$ and $\bar{G}$ are disjoint and B is normal there exist disjoint open sets U and V in B containing $\bar{F}$ and $\bar{G}$. Let U' and V' be the intersections of U and V with A. Then U' and V' are disjoint open sets in A containing F and G.

We can now use Lemma 10 to apply the argument of the paragraph immediately above to R' and S, with R' taking the place of A and S the place of B. Hence R', and its homeomorph R, are normal. This proves Lemma 12.

Collecting these results, we see that we have proved the results announced in the note [7]:

Theorem 1. *Given a distributive lattice L with zero and unit, there is a bicompact T_1-space S, a basis for whose closed sets is a lattice-homeomorphic image of L.*

Corollary. *This basis for the closed sets of S is isomorphic to L if and only if L has the property that if a and b are different elements of L there exists an element c of L such that one of ac and bc is zero and the other is not zero.*

---

[11] Tychonoff [6] proved that a completely regular space could be imbedded in a bicompact Hausdorff space. The present results show that any T_1-space can be imbedded in a bicompact space of the same dimension and that a normal space can be imbedded in a bicompact Hausdorff space of the same dimension. There is still this question left open: can a completely regular space be imbedded in a bicompact Hausdorff space *of the same dimension?*

[12] In terms of the structure of the lattice L of the closed sets of R we could require: if f and g are elements of L and $fg = 0$ then there exist f' and g' in L such that $f' + g' = 1$, $ff' = gg' = 0$. This is, of course, the formulation of normality in terms of closed sets only.

THEOREM 2. *If R is a T_1-space then the bicompact T_1-space S obtained by applying the process of Theorem 1 to the lattice of the closed sets of R is such that S contains a dense subset R' homeomorphic to R; if f and g are closed disjoint subsets of R, and F and G their correspondents in S under the homeomorphism $R \rightleftarrows R'$, $\bar{F}$ and $\bar{G}$ the closures in S of F and G, then $\bar{F}$ and $\bar{G}$ are also disjoint; in the sense of Čech, the homology theory of R is identical with that of S and dimension R = dimension S.*

SPECIAL CASE. *S is a Hausdorff space (and hence normal) if and only if R is a normal space.*

Theorem 1 is proved by Lemmas 4, 5, 6, and 7. The corollary is proved by Lemma 3. Theorem 2 is proved by Lemmas 8, 9, 10, and 11. The special case is considered in Lemma 12.

II. ABSTRACT COMPLEXES AND HOMOLOGY THEORY

15. An *abstract complex* K is an aggregate of certain distinguished finite subsets (repetitions and order being disregarded) of the objects of a finite or infinite collection k, with the sole condition that every subset of a distinguished set is also distinguished. The objects in k are called *vertices*, the distinguished sets are called *simplexes*, and the subsets of a simplex, including the simplex itself, are called the *faces* of a simplex. The *star* of a simplex is the set of all simplexes having the given simplex as face, and the *closed star* is the star together with its faces. As usual no distinction will be made between the simplexes consisting of single vertices and the vertices themselves.

DEFINITION. An abstract complex is *multiplicative* if for every simplex σ there is a vertex a such that the closed star of σ is identical with the closed star of a.

16. DEFINITION. A collection of vertices of K is called a *covering* if every simplex of K is contained in the closed star of some vertex of the covering.

17. DEFINITION. Let a and b be vertices. We say $a < b$ if the closed star of a is contained in the closed star of b.

REMARK. It is evident that $a < a$, and if $a < b$ and $b < c$ then $a < c$.

18. DEFINITION. A collection $\mathfrak{S}$ of vertices which has first the *simplex property*: "any finite subset of $\mathfrak{S}$ is a simplex", and which is second a proper sub-collection of no collection with the simplex property, is called a *maximal collection*, or *point*. The vertices of the collection are called the *coördinates* of the point. The set of all points is denoted by S.

LEMMA 13. *If $\mathfrak{F}$ is any collection of vertices with the simplex property there is a point whose set of coördinates includes $\mathfrak{F}$.*

PROOF. The proof is precisely as in Lemma 1. Note the remark of the last paragraph of §3.

19. Definition. With each vertex a we associate a set of points of S, called the *sub-basic a-set*, consisting of all points having a as a coördinate.

Remark 1. A point belongs to each of the sub-basic sets determined by its coördinates, and since a point is a maximal collection it is the only point common to these sub-basic sets, i.e., *a point is the intersection of the sub-basic sets determined by the coördinates of the point.*

Remark 2. If the point P is not in the sub-basic a-set there is a finite set of coördinates of P such that if τ is the simplex they form, $a\tau$ is not a simplex of K. For otherwise P would not be a maximal collection.

Lemma 14. *Different vertices of k determine different sub-basic sets of S if and only if k has the disjunction property: "$a < b$ and $b < a$ imply $a = b$."*

Proof. Suppose k has the disjunction property. Let a and b be different vertices. Then, we may assume, as a matter of notation, that $b < a$ is false, i.e., there exists a simplex $bc_1 \cdots c_n$ such that $abc_1 \cdots c_n$ is not a simplex. Hence by Lemma 13 there is a point P whose coördinates include the set $b, c_1, \cdots, c_n$; so that P is in the sub-basic b-set. P however is not in the sub-basic a-set since $abc_1 \cdots c_n$ is not a simplex. Thus the sub-basic a-set and the sub-basic b-set are different.

Conversely, suppose that the sub-basic a-set and the sub-basic b-set are different. Then there is a point P which is, as a matter of notation, in the sub-basic b-set but not in the sub-basic a-set. By Remark 2 of this § there is a finite set of coördinates of P, $c_1, \cdots, c_m$ such that $ac_1 \cdots c_m$ is not a simplex. A fortiori, $abc_1 \cdots c_m$ is not a simplex. However, $bc_1 \cdots c_m$ is a simplex since it is a finite subset of the coördinates of P. But this shows that $b < a$ is false, and hence that k has the disjunction property.

20. Lemma 15. *The necessary and sufficient condition that a finite set of vertices of k form a simplex is that the intersection of the sub-basic sets in S associated with them be non-vacuous.*

Proof. If a finite set of vertices of k form a simplex then by Lemma 13 there exists a point of S having these vertices as coördinates, so that the intersection of the sub-basic sets associated with these vertices is non-zero. Conversely, if a finite set of sub-basic sets have a point of S in common the vertices of k with which the sub-basic sets are associated are coördinates of this point, and since every finite set of coördinates of a point forms a simplex, these vertices form a simplex.

Lemma 16. *The necessary and sufficient condition that a finite set of vertices of k be a covering of K is that the sum of the sub-basic sets of S associated with them be equal to S.*

Proof. Let $b_1, \cdots, b_m$ be a covering of K. Suppose there exists a point P of S which is in none of the sub-basic b_i-sets, $i = 1, \cdots, m$. We know by Remark 2 of §19 that for each i there is a finite set of coördinates of P, $a_1^i, a_2^i, \cdots, a_{n(i)}^i$, depending on i, such that $b_i a_1^i a_2^i \cdots a_{n(i)}^i$ is not a simplex. Collecting these vertices for all i we obtain a finite set of coördinates of P, not

depending on i, which we may call $c_1, \cdots, c_r$ such that for each i, $b_i c_1 \cdots c_r$ is not a simplex. But $c_1 \cdots c_r$ is a simplex, since it is a finite set of coördinates of a point. Thus we have encountered a contradiction to the statement that $b_1, \cdots, b_m$ is a covering of S. Hence every point of S is contained in one of the sub-basic b_i-sets.

Conversely let the sum of the sub-basic b_i-sets, $i = 1, \cdots, m$, be equal to S. Let $a_1 \cdots a_n$ be any simplex of K. By Lemma 13 there exists a point P having $a_1, \cdots, a_n$ as coördinates, and since the sub-basic b_i-sets cover S, P also has as coördinate some one, b_{i_0} say, of the vertices b_i. $a_1, \cdots, a_n$ together with b_{i_0} therefore form a finite set of coördinates of P and hence form a simplex of K. This proves that the collection of vertices b_i constitutes a covering of K.

Lemma 17. *The necessary and sufficient condition that $a < b$ is that the sub-basic a-set be contained in the sub-basic b-set.*

Proof. If $a < b$ and P is in the sub-basic a-set then P is in the sub-basic b-set; for otherwise by Remark 2 of §19 there would exist a finite set of coördinates of P (and hence a simplex), which we can take to contain a, such that the subset of k formed by adjoining b is not a simplex, and this contradicts the definition of $a < b$. Conversely let the sub-basic a-set be contained in the sub-basic b-set. Let $ac_1 \cdots c_m$ be any simplex containing a. By Lemma 13 there exists a point P having $a, c_1, \cdots, c_m$ as coördinates. Since P is in the sub-basic a-set and therefore in the sub-basic b-set, P has $a, b, c_1, \cdots, c_m$ as coördinates and these vertices form a simplex, proving that $a < b$.

Lemma 18. *If the intersection of a finite number of sub-basic sets is non-vacuous this intersection is itself a sub-basic set.*

Proof. If $a_1, \cdots, a_k$ are a finite set of vertices of K such that the intersection of the sub-basic a_i-sets is non-vacuous we know from Lemma 15 that $a_1, \cdots, a_k$ form a simplex σ of K. Let a be the vertex of K associated with σ by the definition in §15 of a multiplicative complex. Then the sub-basic a-set = the intersection of the sub-basic a_i-sets.

For since closed star a = closed star $\sigma \subset$ closed star a_i, we have that $a < a_i$. Hence by Lemma 17 the sub-basic a-set $\subset$ sub-basic a_i-set. Conversely if P is a point not in the sub-basic a-set there is by Remark 2 of §19 a simplex τ made up of coördinates of P such that $a\tau$ is not a simplex. Since closed star a = closed star σ, $\sigma\tau$ is not a simplex also. Hence P cannot have as coördinates the vertices a_i and therefore P is not contained in the intersection of the sub-basic a_i-sets.

21. Definition. A *basic* set of S is in the sum of a finite number of sub-basic sets.

Remark. Evidently the sum of two basic sets is a basic set.

Definition. A *closed* set of S is the intersection of a finite or infinite number of basic sets, or else the zero set, or else the unit set.

Lemma 19. *The closed sets satisfy the conditions*:

1) *The zero set and the unit set are closed.*

2) *The sum of two closed sets is closed.*

3) *The intersection of a finite or infinite number of closed sets is closed.*

4) *A point is a closed set.*

PROOF. 1) follows from the definition. 3) is obvious. 4) follows from Remark 1 of §19. If either or both of A and B is the zero set or the unit set, 2) is obvious. To prove 2) in the other cases let A and B be closed sets, $A = \prod A_i$, $B = \prod B_j$, where the A_i and the B_j are basic sets. Then $A + B = \prod (A_i + B_j)$ and since $A_i + B_j$ is a basic set by the Remark of this §, so is $A + B$.

22. We have thus turned the set S into a T_1-space by using the sub-basic sets as a *sub-basis* for the closed sets of the space, i.e., as a collection of closed sets such that every closed set is an infinite intersection of finite sums of these sub-basic closed sets.

LEMMA 20. *S is a bicompact T_1-space.*

PROOF. Let $\mathfrak{A}$ be a collection of closed sets of S any finite number of which have a non-zero intersection. To prove that S is bicompact we have to display a point P of S common to all the closed sets of $\mathfrak{A}$. Replace each closed set of $\mathfrak{A}$ by the finite or infinite collection of basic sets of which it is the intersection. We then have a collection $\mathfrak{B}$ of basic sets, and it is evident that $\mathfrak{B}$ also has the property that any finite number of its elements have a non-zero intersection. Now replace each element of $\mathfrak{B}$ by the finite collection of sub-basic sets of which it is the sum. We obtain a collection $\mathfrak{C}$ of sub-basic sets. The collection $\mathfrak{C}$, however, need not have the property that any finite number of its elements have a non-zero intersection. Nevertheless, we can prove the

LEMMA 21. *Let $B^1, \cdots, B^\alpha, \cdots$ be a collection of sets with the property that any finite number have a non-zero intersection. Decompose each B^α into a finite sum of sets, so that for each α we have $B^\alpha = C_1^\alpha + C_2^\alpha + \cdots + C_m^\alpha(\alpha)$, $m(\alpha)$ being finite. Then it is possible to select for each α an integer $i(\alpha)$ out of the range $1, \cdots, m(\alpha)$ such that the collection $\{C_{i(\alpha)}^\alpha\}$ has the property that any finite number of its sets have a non-zero intersection.*

PROOF OF LEMMA 21. We shall show by a transfinite induction (we assume that the B's are well-ordered) that *it is possible to select for each α an integer $i(\alpha)$ out of the range $1, \cdots, m(\alpha)$ such that if $\alpha_1, \cdots, \alpha_r$ is any finite set of ordinal numbers each $\leqq \alpha$ and if $\alpha_1', \cdots, \alpha_s'$ is any finite set of ordinal numbers each $> \alpha$, then the sets $C_{i(\alpha_1)}^{\alpha_1}, \cdots, C_{i(\alpha_r)}^{\alpha_r}, C_{i(\alpha)}^{\alpha}, B^{\alpha_1'}, \cdots, B^{\alpha_s'}$ have a non-zero intersection.* The sets $C_{i(\alpha)}^\alpha$ will then evidently fulfill the requirements of the lemma.

Starting the induction with $\alpha = 1$, we see that there exists an integer $i(1)$ out of the range $1, \cdots, m(1)$ such that for every finite number of B^β, $\beta > 1$, the intersection of $C_{i(1)}^1$ and these B^β is not zero. Otherwise for each $j = 1, \cdots, m(1)$ there would exist a finite set of ordinal numbers depending on j, $1'(j), \cdots, n_j'(j)$, each $k'(j) > 1$, such that $C_j^1 \cdot B^{1'(j)} \cdots B^{n_j'(j)} = 0$. If we collect together the $k'(j)$ for all j and write their set as $l'(1), \cdots, l'(q)$, we have

a finite set of ordinal numbers, not depending on j, with each $l' > 1$, such that $C_j^1 \cdot B^{l'(1)} \cdots B^{l'(q)} = 0$ for each $j = 1, \cdots, m(1)$. Hence

$$(C_1^1 + C_2^1 + \cdots + C_{m(1)}^1)B^{l'(1)} \cdots B^{l'(q)} = 0.$$

But $C_1^1 + \cdots + C_{m(1)}^1 = B^1$ so that $B^1B^{l'(1)} \cdots B^{l'(q)} = 0$. This however contradicts the hypothesis that every finite number of B's have a non-zero intersection. Hence an $i(1)$ exists satisfying the condition of the induction.

Now let us assume the assertion of the induction for all $\beta < \alpha$. Then the assertion follows for $\beta = \alpha$. Otherwise for each $j = 1, \cdots, m(\alpha)$ there exist two finite sets of ordinal numbers depending on j, $1(j_\alpha), \cdots, n_j(j_\alpha)$, each $k(j_\alpha) \leqq \alpha$, and $1'(j_\alpha), \cdots, n_j'(j_\alpha)$, each $k'(j_\alpha) > \alpha$, such that

$$C_{i[1(j_\alpha)]}^{1(j_\alpha)} \cdots C_{i[n_j(j_\alpha)]}^{n_j(j_\alpha)} C_{i(\alpha)}^{\alpha} B^{1'(j_\alpha)} \cdots B^{n_j'(j_\alpha)} = 0.$$

If we collect together the $k(j_\alpha)$ and the $k'(j_\alpha)$ for all j and write their sets as $l_\alpha(1), \cdots, l_\alpha(p)$ and $l_\alpha'(1), \cdots, l_\alpha'(q)$ respectively, we have two finite sets of ordinal numbers, not depending on j, with each $l_\alpha \leqq \alpha$ and each $l_\alpha' > \alpha$, such that

$$C_{i[l_\alpha(1)]}^{l_\alpha(1)} \cdots C_{i[l_\alpha(p)]}^{l_\alpha(p)} C_{i(\alpha)}^{\alpha} B^{l_\alpha'(1)} \cdots B^{l_\alpha'(q)} = 0$$

for each $j = 1, \cdots, m(\alpha)$. Hence

$$C_{i[l_\alpha(1)]}^{l_\alpha(1)} \cdots C_{i[l_\alpha(p)]}^{l_\alpha(p)} (C_1^\alpha + \cdots + C_{m(\alpha)}^\alpha) B^{l_\alpha'(1)} \cdots B^{l_\alpha'(q)} = 0.$$

Now $C_1^\alpha + \cdots + C_{m(\alpha)}^\alpha = B^\alpha$, so that

$$C_{i[l_\alpha(1)]}^{l_\alpha(1)} \cdots C_{i[l_\alpha(p)]}^{l_\alpha(p)} B^\alpha B^{l_\alpha'(1)} \cdots B^{l_\alpha'(q)} = 0.$$

But this contradicts the hypothesis of the induction. Hence there does exist an integer $i(\alpha)$ out of the range $1, \cdots, m(\alpha)$ for which the assertion is true.

This proves Lemma 21.

Returning now to the proof of Lemma 20 we see that we can apply Lemma 21 to determine a collection $\mathfrak{C}'$ of sub-basic sets which is a subcollection of $\mathfrak{C}$, and which is such that each basic set of $\mathfrak{B}$ has one of the sub-basic sets of which it is the sum as an element in $\mathfrak{C}'$, and such that any finite number of elements of $\mathfrak{C}'$ have a non-zero intersection. Associated with $\mathfrak{C}'$ is the set $\mathfrak{D}$ of all vertices of k whose corresponding sub-basic sets are in $\mathfrak{C}'$. Then from Lemma 15 we know that $\mathfrak{D}$ has the property that any finite number of its vertices form a simplex. Hence by Lemma 13 there exists a point P whose set of coördinates includes $\mathfrak{D}$; P is therefore contained in the intersection of the sub-basic sets $\mathfrak{C}'$, hence in the intersection of the basic sets $\mathfrak{B}$, and hence finally in the intersection of the sets $\mathfrak{A}$, q.e.d.

23. Let us say that two finite coverings by closed sets of S are *equivalent* if there is a 1:1 correspondence between their elements which preserves the property of having a non-zero intersection.

Lemma 22. *If S is a Hausdorff space, then given any finite covering by closed*

sets of S there is an equivalent finite covering by closed sets which has a refinement made up of sub-basic sets.

Proof. α) Given any (additive) basis $\mathfrak{A}$ for the open sets of a bicompact Hausdorff space S and any finite covering by closed sets $\mathfrak{U}$: $F_1, \cdots, F_k$ of S there is an equivalent finite covering by closed sets $\mathfrak{V}$: $G_1, \cdots, G_k$, with each G_j a sum of closures of sets in $\mathfrak{A}$.

For, given $\mathfrak{A}$ there is[13] a collection of open sets $U_1, \cdots, U_k$ such that $F_j \subset U_j$ and $U_{i_0} \cdots U_{i_n} \neq 0$ is equivalent to $F_{i_0} \cdots F_{i_n} \neq 0$. Since S is normal there exists for each j, $j = 1, \cdots, k$, an open set V_j such that $F_j \subset V_j$ and $\bar{V}_j \subset U_j$. Since V_j is open it is the sum of sets of the open basis $\mathfrak{A}$. Since F_j is closed and S is bicompact, F_j is covered by a finite number $u_j^1, \cdots, u_j^{m(j)}$, $m(j)$ finite, of the basic open sets. Let $\bar{u}_j^i$ be the closure of u_j^i. Let $G_j = \sum_{i=1}^{m(j)} \bar{u}_j^i$. Then $F_j \subset G_j \subset \bar{V}_j$, so that $F_j \subset G_j \subset U_j$. Hence $G_{i_0} \cdots G_{i_n} \neq 0$ is equivalent to $F_{i_0} \cdots F_{i_n} \neq 0$. This proves α).

β) Let $\mathfrak{A}$ be any (additive) basis for the open sets of a bicompact Hausdorff space S. Then the closures of the sets in $\mathfrak{A}$ form a sub-basis for the closed sets of S.

For, it follows simply from the concept of normality that given any closed set F of S there is an infinite collection of open sets U^α such that $F \subset U^\alpha$ and $\prod \bar{U}^\alpha = F$. Since U^α is open it is a sum of sets in $\mathfrak{A}$, and since F is closed and S is bicompact, F is covered by a finite set $u_1, \cdots, u_{m(\alpha)}$, $m(\alpha)$ finite, of these basic open sets. Since $F \subset \sum_{i=1}^{m(\alpha)} \bar{u}_i^\alpha \subset \bar{U}^\alpha$, we have that $\prod_\alpha \sum_{i=1}^{m(\alpha)} \bar{u}_i^\alpha = F$, which proves β).

γ) Given any (multiplicative) basis $\mathfrak{M}$ for the closed sets of a bicompact Hausdorff space S, and a finite covering by closed sets $\mathfrak{U}$: $F_1, \cdots, F_k$ of S there is an equivalent finite covering by closed sets $\mathfrak{V}$: $G_1, \cdots, G_k$ with each G_j a sum of finite intersections of sets in $\mathfrak{M}$.

For, using β) and replacing each set by its complement, we see that the interiors of the sets of $\mathfrak{M}$ form an (open) sub-basis for the open sets of S, i.e. finite intersections of the interiors of sets in $\mathfrak{M}$ form an (additive) basis for the open sets of S. Let this open basis be the $\mathfrak{A}$ of α). Then applying α) the statement of γ) is immediate.

δ) Given any sub-basis $\mathfrak{B}$ for the closed sets of a bicompact Hausdorff space, and a finite covering by closed sets $\mathfrak{U}$: $F_1, \cdots, F_k$ of S there is an equivalent finite covering by closed sets $\mathfrak{V}$: $G_1, \cdots, G_k$ with each G_j a sum of finite intersections of sets in $\mathfrak{B}$.

For, let the $\mathfrak{M}$ in γ) be the basis for closed sets of S whose elements consist of finite sums of the sets in $\mathfrak{B}$. Each summand of a G_j is by γ) a finite intersection of sets in $\mathfrak{M}$, and applying the distributive law, is therefore a sum of finite intersections of sets in $\mathfrak{B}$. This proves δ).

ϵ) For the particular sub-basis of S consisting of the sub-basic sets associated with the vertices of K, the statement of Lemma 22 is true.

---

[13] [3], p. 179. Here we use the fact that S is normal (since it is bicompact and Hausdorff).

For, let us apply Lemma 18 to δ), and let us replace each G_i by the collection of sub-basic sets of which it is the sum.

This proves Lemma 22.

Combining the results of Lemmas 14, 15, 16, 17, and 20 we obtain

THEOREM 3. *Given any (infinite) abstract complex K which has the disjunction property (see Lemma 14), there is a bicompact T_1-space S and a sub-basis for the closed sets of S, such that the nerve K' of the aggregate of the sets of the sub-basis is isomorphic to K, in the sense that the properties of forming a simplex, of forming a covering, and of being less than, are preserved in the 1:1 correspondence between K and K'.*

It is evident that Lemma 22 and Theorem 3 prove

THEOREM 4. *If S is a Hausdorff space the homology theory of the abstract multiplicative complex K is identical with that of the space S.*

PRINCETON UNIVERSITY.

BIBLIOGRAPHY

[1] P. ALEXANDROFF, *Zur Homologietheorie der Kompakten*, Comp. Math., v. 4 (1937).
[2] P. ALEXANDROFF-H. HOPF, *Topologie* I, Julius Springer, Berlin (1935).
[3] E. ČECH, *Théorie générale de l'homologie dans un espace quelconque*, Fund. Math., t. 19 (1932).
[4] A. KUROSCH, *Kombinatorischer Aufbau bikompakter Räume*, Comp. Math., v. 2 (1935).
[5] M. H. STONE, *Applications of the Theory of Boolean Rings to General Topology*, Trans. Am. Math. Soc., v. 41 (1937).
[6] A. TYCHONOFF, *Über die topologische Erweiterung von Räumen*, Math. Ann., Bd. 102 (1930).
[7] H. WALLMAN, *Lattices and Bicompact Spaces*, Proc. Nat. Acad., v. 23 (1937).

ANNALS OF MATHEMATICS
Vol. 39, No. 1, January, 1938

TRANSFORMATIONS OF FINITE PERIOD

BY P. A. SMITH

(Received May 25, 1937)

This paper deals with certain questions connected with the following general problem: to determine the extent to which a periodic homeomorphic transformation of Euclidean n-space into itself resembles a rotation or at least an orthogonal transformation. This formulation is essentially equivalent to the more convenient one obtained by replacing "Euclidean n-space" by "n-sphere." With this in mind, the problem may be considered as being completely solved when the dimension n is two. For it was shown by Brouwer [12] and independently by Kérékjártó [13] that every periodic transformation of an ordinary sphere M into itself is topologically equivalent to a rotation, or to the product by a rotation by a reflection across a diametral plane. For $n > 2$, the difficulties which make a similar result seem unattainable at the present time are well known to topologists. But if it is necessary to abandon the idea of determining completely the structure of periodic transformations it may nevertheless be of interest to study such of their properties as can be described in terms of homology theory. Consider for example the set L of fixed points and assume that $L \neq 0$. Under an orthogonal transformation of M the set L must itself be a sphere of dimension $\leqq n$. Does this property of L hold for an arbitrary periodic transformation T? We shall show that the property in question—at least if the order p of T is prime—does hold in the sense that L has the same Betti numbers modulo p, both locally and in the large, as an r-sphere ($0 \leqq r \leqq n - 1$); moreover the "modulo p dimension" (Alexandroff) of L is r.

It is of course to be expected that results of this type should not depend so much upon the fact that M is actually a sphere as on the fact that M has such and such homology characters. This is indeed the case: we need assume little more concerning M than that it be a bicompact space having the same homology groups (in the sense of Čech) as an n-sphere. It is perhaps interesting to remark that due to a theorem of Alexandroff, the mere addition of the assumption that M be compact and metric yields the corollary that if the Menger-Urysohn dimension of L is one, then L is a simple closed curve. Among various other special results we shall only mention (1) the almost complete analysis of the structure of L when M is a 3-sphere; (2) the evaluation of certain homology characters of the fundamental domain M^* of M relative to any periodic T. M^* is defined as the space obtained form M by identifying points which are images of one another under powers of T. In a sense, the structure of M^* may be considered as characteristic of the structure of T itself.[1]

[1] Certain of our results were first announced in 1935 at the international topological congress at Moscow; see [14].

1. Preliminary Theorems

Throughout this paper we shall use the form of homology theory that was developed by E. Čech in his fundamental paper [1]. We shall therefore take for granted a certain familiarity with the Čech theory, particularly with the properties of normal refinements and essential cycles.

1.1. *Let M be a completely normal space* ([11], p. 265) *and let L, K be closed sets in M. Let $\{\mathfrak{U}\}$ be the totality of finite coverings $\mathfrak{U}$ of M by open sets and let $X_h = \{X_h(\mathfrak{U})\}$ be an h-cycle* mod $(K + L)$ *(in M) with coefficients in a field* $\mathfrak{f}$. *There exists a cycle X_h'* mod $(K + L)$ *and a cycle γ_{h-1}'* mod KL *in L such that $X_h' \sim X_h$* mod $(K + L)$ *and $\phi X_h' = \gamma_{h-1}'$* mod K.[2]

Proof. For each $\mathfrak{U}$ there exists a refinement $\mathfrak{U}_1$ with the property that every $\mathfrak{U}_1$-chain which is simultaneously in L and in K is in KL ([5], p. 8). Let a definite $\mathfrak{U}_1$ be chosen for every $\mathfrak{U}$. Let $\gamma_{h-1}(\mathfrak{U}_1) = L \cap \phi X_h(\mathfrak{U}_1)$—that is, let $\gamma_{h-1}(\mathfrak{U}_1)$ be the subchain of $\phi X_h(\mathfrak{U}_1)$ which consists of those simplexes of $\phi X_h(\mathfrak{U}_1)$ which are in L, taken with the same coefficients that they have in $\phi X_h(\mathfrak{U}_1)$. $\phi\gamma_{h-1}(\mathfrak{U}_1) \subset L$ since $\gamma_{h-1}(\mathfrak{U}_1) \subset L$. Moreover the chain $D_{h-1}(\mathfrak{U}_1) = \phi X_h(\mathfrak{U}_1) - \gamma_{h-1}(\mathfrak{U}_1)$ is in K so that $\phi X_h(\mathfrak{U}_1) = \gamma_{h-1}(\mathfrak{U}_1)$ mod K. Since $\phi\phi X_h(\mathfrak{U}_1) = 0$, $\phi D_{h-1}(\mathfrak{U}_1) = -\phi\gamma(\mathfrak{U}_1)$ so that $\phi\gamma_{h-1}(\mathfrak{U}_1)$ is also in K. Hence $\phi\gamma_{h-1}(\mathfrak{U}_1) \subset KL$ and hence $\gamma_{h-1}(\mathfrak{U}_1)$ is a $\mathfrak{U}_1$-cycle mod KL in L. We shall show that relative to the complete family $\{\mathfrak{U}_1\}$([1] p. 165), $\gamma_{h-1} = \{\gamma_{h-1}(\mathfrak{U}_1)\}$ is a cycle mod KL in L. To verify this we must show that if $\mathfrak{B}_1 \subset \mathfrak{U}_1$ (i.e. if $\mathfrak{B}_1$ is a refinement of $\mathfrak{U}_1$) and if $\pi = \pi(\mathfrak{B}_1, \mathfrak{U}_1) =$ a projection of $\mathfrak{B}_1$ into $\mathfrak{U}_1$, then $\pi\gamma_{h-1}(\mathfrak{B}_1) \sim \gamma_{h-1}(\mathfrak{U}_1)$ mod KL in L. We have $\pi X(\mathfrak{B}_1) \sim X(\mathfrak{U}_1)$ mod $(K + L)$ which means that there exist chains $Y_{h-1}(\mathfrak{U}_1)$ and $H_h(\mathfrak{U}_1)$, the latter in $K + L$, such that

$$\phi Y_{h+1}(\mathfrak{U}_1) = \pi X(\mathfrak{B}_1) - X(\mathfrak{U}_1) - H(\mathfrak{U}_1).$$

The expression on the right is an absolute cycle;[3] hence if we equate its boundary to zero, and recall that π and ϕ are permutable, we obtain

$$(1) \qquad \phi H(\mathfrak{U}_1) = \pi(\gamma(\mathfrak{B}_1) + D(\mathfrak{B}_1)) - \gamma(\mathfrak{U}_1) - D(\mathfrak{U}_1) \sim 0.$$

Let $H'(\mathfrak{U}_1) = L \cap H(\mathfrak{U}_1)$, $H''(\mathfrak{U}_1) = H(\mathfrak{U}_1) - H'(\mathfrak{U}_1)$. Then $H'(\mathfrak{U}_1) \subset L$, $H''(\mathfrak{U}_1) \subset K$ and (1) can be written the form

$$\phi H'(\mathfrak{U}_1) = \pi\gamma(\mathfrak{B}_1) - \gamma(\mathfrak{U}_1) + J(\mathfrak{U}_1)$$

where $J(\mathfrak{U}_1) \subset K$ (since $\pi D(\mathfrak{B}_1)$ and $D(\mathfrak{U}_1)$ are in K). But since $\phi H'(\mathfrak{U}_1)$, $\pi\gamma(\mathfrak{B}_1)$, $\gamma(\mathfrak{U}_1)$ are in L, $J(\mathfrak{U}_1)$ is also in L, hence in KL. Hence $\pi\gamma(\mathfrak{B}_1) \sim \gamma(\mathfrak{U}_1)$ mod KL in L.

Let a projection $\pi_1 = \pi_1(\mathfrak{U}_1, \mathfrak{U})$ be chosen for every $\mathfrak{U}$ and let $X'(\mathfrak{U}) =$

---

[2] i.e. $\phi X_h'(\mathfrak{U}) = \gamma_{h-1}'(\mathfrak{U})$ mod K for every $\mathfrak{U}$, where ϕ means "the boundary of . . ."

[3] i.e. an $(h, \mathfrak{U}_1)$—cycle. We shall frequently omit symbols which represent the dimension (always written as a subscript), the particular covering being used etc. when the meaning is clear from the context.

$\pi_1 X(\mathfrak{U}_1)$, $\gamma'(\mathfrak{U}) = \pi_1\gamma(\mathfrak{U}_1)$. Since $\{X(\mathfrak{U})\}$ is a cycle, we have $X'(\mathfrak{U}) \sim X(\mathfrak{U})$ mod $(K + L)$. Hence $X' = \{X'(\mathfrak{U})\}$ is a cycle mod $(K + L)$ and $X' \sim X$ mod $(K + L)$. Moreover, each $\gamma'(\mathfrak{U})$ is a $\mathfrak{U}$-cycle mod KL in L and it is easy to show that $\gamma' = \{\gamma'(\mathfrak{U})\}$ is a cycle mod KL in L (cf. [1] p. 165). Since the chain $\phi X(\mathfrak{U}_1) - \gamma(\mathfrak{U}_1)$ equals zero mod K, so does its projection by π_1. Hence $\phi X' = \gamma'$ mod K, which completes the proof.

1.2. *Let L be a set in M*[4] *and let X_h be a cycle* mod L. *There exists a cycle γ_{h-1} in L such that $\phi X_h = \gamma_{h-1}$.*

PROOF. It can be immediately verified that $\{\phi X_h(\mathfrak{U})\}$ is a cycle in L.

1.3. *Let K, L be subsets of M and let γ_{h-1} be a cycle* mod KL *in L and X_h a cycle* mod $(K + L)$ *such that $\phi X = \gamma$* mod K. *If $\gamma \sim 0$* mod KL *in L and if every h-cycle* mod K *is ~ 0* mod K, *then $X \sim 0$* mod $(K + L)$.

PROOF. By hypothesis, there exists for each $\mathfrak{U}$ a chain $Y_h(\mathfrak{U})$ in L such that $\phi Y(\mathfrak{U}) = \gamma_{h-1}(\mathfrak{U})$ mod KL. This relation also holds mod K. Since $\phi X(\mathfrak{U}) = \gamma_{h-1}(\mathfrak{U})$ mod K, it follows that $Y(\mathfrak{U}) - X(\mathfrak{U})$ is a cycle mod K, hence mod $(K + L)$. Let $\mathfrak{U}_1$ be a normal refinement of $\mathfrak{U}$ relative to cycles mod $(K + L)$ and let $X'(\mathfrak{U}) = \pi_1 X(\mathfrak{U}_1)$, $Y'(\mathfrak{U}) = \pi_1 Y(\mathfrak{U}_1)$, $\pi_1 = \pi_1(\mathfrak{U}_1, \mathfrak{U})$. Then $X'(\mathfrak{U}) \sim X(\mathfrak{U})$ mod $(K + L)$ and the cycle $X'(\mathfrak{U}) - Y'(\mathfrak{U})$ is essential and therefore ~ 0 mod $(K + L)$.[4a] Therefore, since $Y'(\mathfrak{U}) \subset L$, we have $X(\mathfrak{U}) \sim X'(\mathfrak{U}) \sim 0$ mod $(K + L)$.

1.4. *Let K, L be closed sets in the completely normal space M. Let γ_{h-1} be a cycle* mod KL *in L and X_h a cycle* mod $(K + L)$ *such that $\phi(X_h) = \gamma_{h-1}$* mod K. *If $X_h \sim 0$* mod $(K + L)$, *then $\gamma_{h-1} \sim 0$* mod KL *in L.*

PROOF. The relation $X_h \sim 0$ mod $(K + L)$ implies that there exists for every $\mathfrak{U}$ a $Y_{h+1}(\mathfrak{U})$ such that $\phi Y_{h+1}(\mathfrak{U}) = X_h(\mathfrak{U}) + Z_h(\mathfrak{U})$ where $Z_h(\mathfrak{U}) \subset K + L$. Let $Z'_h(\mathfrak{U}) = L \cap Z_h(\mathfrak{U})$ and $Z''(\mathfrak{U}) = Z_h(\mathfrak{U}) - Z'_h(\mathfrak{U})$. Then $Z''_h(\mathfrak{U}) \subset K$. We have now

$$(1) \qquad 0 = \phi\phi Y_{h+1}(\mathfrak{U}) = \phi X_h(\mathfrak{U}) + \phi Z'_h(\mathfrak{U}) + \phi Z''_h(\mathfrak{U}),$$

and since $\phi X_h(\mathfrak{U}) = \gamma_{h-1}(\mathfrak{U}) + G_{h-1}(\mathfrak{U})$ where $G_{h-1}(\mathfrak{U}) \subset K$, the relation (1) can be written

$$(2) \qquad \phi Z'_h(\mathfrak{U}) = \gamma_{h-1}(\mathfrak{U}) - H_{h-1}(\mathfrak{U})$$

where $H_{h-1}(\mathfrak{U}) \subset K$. Now since $Z'_{h+1}(\mathfrak{U}) \subset L$ and $\gamma_h(\mathfrak{U}) \subset L$, we have $H_{h-1}(\mathfrak{U}) \subset L$. Let $\mathfrak{U}_1$ be a refinement of $\mathfrak{U}$ such that chains which are simulta-

---

[4] In 1.2, 1.3, and 1.6 M may be any abstract set and $\{\mathfrak{U}\}$ any fundamental family of coverings of M. In 1.4 and 1.5 it is to be understood that $\{\mathfrak{U}\}$ is the same as in 1.1.

[4a] For, there exists a (Čech) cycle Z_h mod $(K + L)$ of which $X'(\mathfrak{U}) - Y'(\mathfrak{U})$ is a "coefficient" ([1]), p. 168). Since by hypothesis $Z \sim 0$ mod $(K + L)$, the same relation holds for each coefficient of Z.

neously in K and in L are in KL (cf. proof of 1.1). Since the covering $\mathfrak{U}$ is arbitrary, all relations written above hold for $\mathfrak{U}_1$. In particular $H(\mathfrak{U}_1)$ is in K and L, hence in KL. Let $\pi_1 = \pi_1(\mathfrak{U}_1, \mathfrak{U})$. Then from (2) we have on projecting by π_1:

$$\pi_1\phi Z'(\mathfrak{U}_1) = \pi_1\gamma_{h-1}(\mathfrak{U}_1) + \pi_1 H_{h-1}(\mathfrak{U}_1). \tag{3}$$

Since $\pi_1\gamma_{h-1}(\mathfrak{U}_1) \sim \gamma_{h-1}(\mathfrak{U})$ mod KL in L, and since $\pi_1 Z'_h(\mathfrak{U}_1)$ and $\pi_1 H_{h-1}(\mathfrak{U}_1)$ are in L and KL respectively, it follows that from (3) that $\gamma_{h-1}(\mathfrak{U}) \sim 0$ mod KL in L.

1.5. *Let K, L be closed subsets of the completely normal space M and let γ_{h-1} be a cycle* mod KL *in L. If $\gamma_{h-1} \sim 0$* mod KL *and if there exists a closed set $K_1 \supset K$ such that every h-cycle* mod K *is ~ 0* mod K_1, *then there exists a cycle $X_{0,h}$* mod $(K_1 + L)$ *and a cycle $\gamma_{0,h-1}$* mod K_1L *in L such that $\phi X_{0,h} = \gamma_{0,h-1}$* mod $(K_1 + L)$ *and $\gamma_{0,h-1} \sim \gamma_{h-1}$* mod K_1L *in L.*

PROOF. Let $\mathfrak{U}_1$, $\mathfrak{U}_2$ be normal refinements of a given $\mathfrak{U}$ relative to cycles mod K_1. If $\gamma_{h-1} \sim 0$ mod KL, then also $\gamma_{h-1} \sim 0$ mod K_1 and hence there exist chains $X'_h(\mathfrak{U}_i)$ $(i = 1, 2)$ such that $\phi X'_h(\mathfrak{U}_i) = \gamma_h(\mathfrak{U}_i)$ mod K_1. Let $\pi_i = \pi_i(\mathfrak{U}_i, \mathfrak{U})$. The chains $X'_h(\mathfrak{U}_i)$ are cycles mod $(K_1 + L)$. We shall show that

$$\pi_1 X'_h(\mathfrak{U}_1) \sim \pi_2 X'_h(\mathfrak{U}_2) \text{ mod } (K_1 + L). \tag{1}$$

Let $\mathfrak{W} \subset \mathfrak{U}_1\mathfrak{U}_2$, (i.e. let $\mathfrak{W}$ be a common refinement of $\mathfrak{U}_1$, $\mathfrak{U}_2$) and let $\pi_{wi} = \pi_{wi}(\mathfrak{W}, \mathfrak{U}_i)$. Let $X'_h(\mathfrak{W})$ be a chain bounded mod K_1 by γ_{h-1}. Then $\pi_{wi}X'(\mathfrak{W}) - X'_h(\mathfrak{U}_i)$ $(i = 1, 2)$ are $\mathfrak{U}_i$-cycles mod $(K_1 + L)$. Since $\pi_{wi}\gamma(\mathfrak{W}) \sim \gamma(\mathfrak{U}_i)$ mod K_1L in L, we may write $\phi Y_h(\mathfrak{U}_i) = \pi_{wi}\gamma(\mathfrak{W}) - \gamma(\mathfrak{U}_i)$ mod K_1L, when $Y_h(\mathfrak{U}_i) \subset L$. Since $\phi\pi_{wi}X(\mathfrak{W}) = \pi_{wi}\gamma(\mathfrak{W})$ mod K_1 and $\phi X'(\mathfrak{U}_i) = \gamma(\mathfrak{U}_i)$ mod K_1, it follows that the chains $\pi_{wi}X'(\mathfrak{W}) - X'(\mathfrak{U}_i) - Y(\mathfrak{U}_i)$ $(i = 1, 2)$ are cycles mod K_1, hence so are their projections by π_i. The latter cycles are essential and therefore ~ 0 mod K_1, hence so is their difference; that is

$$\pi_1 X'(\mathfrak{U}_1) - \pi_2 X'(\mathfrak{U}_2) - (\pi_1\pi_{w1}X'(\mathfrak{W}) - \pi_2\pi_{w2}X'(\mathfrak{W})) + (\pi_1 Y(\mathfrak{U}_1) - \pi_2 Y(\mathfrak{U}_2)) \sim 0 \text{ mod } K_1.$$

This however reduces to (1) if we observe that $\pi_1\pi_{w1}X'(\mathfrak{W}) \sim \pi_2\pi_{w2}X'(\mathfrak{W})$ mod $(K_1 + L)$ and $\pi_1 Y(\mathfrak{U}_1) - \pi_2 Y(\mathfrak{U}_2) \subset L$.

Now for *each* $\mathfrak{U}$ let there be chosen a normal refinement $\mathfrak{U}_1$ relative to cycles mod $(K_1 + L)$ and let $X(\mathfrak{U}) = \pi_1 X'(\mathfrak{U}_1)$, $\pi_1 = \pi_1(\mathfrak{U}_1, \mathfrak{U})$. We assert that $\{X(\mathfrak{U})\}$ is a cycle mod $(K_1 + L)$. Suppose $\mathfrak{V} \subset \mathfrak{U}$, $\pi = \pi(\mathfrak{V}, \mathfrak{U})$. We must show that $\pi X_h(\mathfrak{V}) \sim X(\mathfrak{U})$ mod $(K_1 + L)$. Let $\mathfrak{X} \subset \mathfrak{U}_1\mathfrak{V}$, $\mathfrak{X}$ normal to $\mathfrak{U}_1$ and $\mathfrak{V}$ relative to cycles mod $(K_1 + L)$. Let $\pi_{xv} = \pi_{xv}(\mathfrak{X}, \mathfrak{V})$. By the last paragraph, $\pi_{xv}X_h(\mathfrak{X}) \sim X_h(\mathfrak{V})$ mod $(K_1 + L)$, and $\pi\pi_{xv}X(\mathfrak{X}) \sim X_h(\mathfrak{U})$ mod $(K_1 + L)$. From the first of these relations, we have $\pi\pi_{xv}X_h(\mathfrak{X}) \sim \pi X_h(\mathfrak{V})$

mod $(K_1 + L)$. Therefore $\pi X_h(\mathfrak{V}) \sim X_h(\mathfrak{U})$ mod $(K_1 + L)$, and X_h is a cycle mod $(K_1 + L)$.

Now by 1.1 there exists a cycle $X'_{0,h}$ mod $(K_1 + L)$ and a cycle $\gamma'_{0,h-1}$ mod K_1L in L such that $X'_{0,h} \sim X_h$ mod $(K_1 + L)$ and $\phi X'_{0,h} = \gamma'_{0,h-1}$ mod K_1. We have then

$$\phi(X'_h(\mathfrak{U}_1) - X'_{0,h}(\mathfrak{U}_1)) = \gamma_{h-1}(\mathfrak{U}_1) - \gamma'_{0,h-1}(\mathfrak{U}_1) \bmod K_1 . \tag{2}$$

Let $\gamma_{0,h-1}(\mathfrak{U}) = \pi_1\gamma'_{0,h-1}(\mathfrak{U}_1)$, $X_{0,h}(\mathfrak{U}) = \pi_1 X'_{0,h}(\mathfrak{U}_1)$ and $\overset{0}{\gamma}_{h-1}(\mathfrak{U}) = \pi_1\gamma_{h-1}(\mathfrak{U}_1)$. Then $X_{0,h} = \{X_{0,h}(\mathfrak{U})\}$ is a cycle mod $(K_1 + L)$ and $X_{0,h} \sim X'_{0,h}$ mod $(K_1 + L)$. Similarly $\overset{0}{\gamma}_{h-1} \sim \gamma_{h-1}$ mod K_1L in L and $\gamma_{0,h-1} \sim \gamma'_{0,h-1}$ mod K_1L in L. On projecting the relation (2) by π_1 into $\mathfrak{U}$ we see that

$$\phi(X_h - X_{0,h}) = \overset{0}{\gamma}_{h-1} - \gamma_{0,h-1} \bmod K_1 .$$

But since $X_h \sim X'_{0,h} \sim X_{0,h}$ mod $(K_1 + L)$ we have $X_h - X_{0,h} \sim 0$ mod $(K_1 + L)$, hence by 1.4 $\gamma_{0,h-1} \sim \overset{0}{\gamma}_{h-1}$ mod K_1L in L. Since, finally, $\phi X_{0,h} = \gamma_{0,h-1}$ mod K_1 the proof of the theorem is complete.

1.6. It will be convenient for later reference to state the following theorem the proof of which is contained essentially in that of 1.5.

Let γ_{h-1} be a cycle in L and let $\{X'_h(\mathfrak{U})\}$ be a family of chains such that $\phi X'_h(\mathfrak{U}) = \gamma_{h-1}(\mathfrak{U})$ for each $\mathfrak{U}$. Assume that all h-cycles are ~ 0. For each $\mathfrak{U}$ let there be chosen a normal refinement $\mathfrak{U}_1$ relative to absolute cycles and let $X_h(\mathfrak{U}) = \pi_1 X'_h(\mathfrak{U}_1)$ $(\pi_1 = \pi_1(\mathfrak{U}_1, \mathfrak{U})$. Then $X_h = \{X_h(\mathfrak{U})\}$ is a cycle mod L.

2. Special Coverings

2.1. We shall now assume that M is a bicompact space. Let p be a prime number and let T be a one-one bi-continuous transformation of M into itself such that T^q is the identity when and only when $q = 0$ (mod p). Let L denote the points which are invariant under T. L is obviously closed, therefore bicompact. *We shall assume that $L \neq 0$.*

Let the images under successive powers of T, of an arbitrary set $A \subset M$, be denoted by $A^0 = A, A^1, A^2, \cdots$ and let $\sigma A = A^0 + A^1 + \cdots + A^{p-1}$. Since p is a prime it is readily seen that when x is an arbitrary point of $M - L$, σx consists of p distinct points and when $x \subset L$, $\sigma x = x$.

2.2. Let x be an arbitrary point of M. Consider now the space M^* whose "points" are the sets σx. Let σx, regarded as a point in M^*, be denoted by x^*. *Two points $x^* = \sigma x$ and $y^* = \sigma y$ of M^* are to be regarded as identical if and only if σx and σy are identical point-sets in M*; obviously $x^* = y^*$ if and only if there exists an integer i $(0 \leq i \leq p - 1)$ such that $x = y^i$.

We now define a single-valued correspondence Λ: $M \to M^*$, by writing $\Lambda x = x^*$ where $x^* = \sigma x$. Obviously $\Lambda x = \Lambda x^1 = \cdots = \Lambda x^{p-1}$. If $A \subset M$, we shall denote by ΛA the set $\{\Lambda x, x \subset A\}$; ΛL will be consistently denoted by L^*.

Let $\Lambda^{-1}A^*(A^* \subset M^*)$ denote the totality of points x such that $\Lambda x \subset A^*$. For arbitrary sets $A^* \subset M^*$ and $A \subset M$, the relations

$$\Lambda^{-1}\Lambda A = \sigma A, \qquad \Lambda\Lambda^{-1}A^* = A^*$$

can be immediately verified. If $x^* \subset L^*$, say $x^* = \Lambda x$, $x \subset L$, then from the first relation we have $\Lambda^{-1}x^* = \sigma x = x$. Thus, over L^* the correspondence Λ^{-1} is single-valued and Λ therefore induces a one-one correspondence between L and L^*.

2.3. Now let the open sets of M^* be defined as being those of the form ΛA where A is an open set in M. *With M^* thus topologised, Λ is continuous.* For let A^* be an open set in M^*, say $A^* = \Lambda A$, A open in M. Then $\Lambda^{-1}A^* = \Lambda^{-1}\Lambda A = \sigma A$ and since σA is open, the continuity of Λ is proved. In particular *Λ induces a homeomorphism between L and L^*.* For if the set $B^* \subset L^*$ is open in L^*, we may write $B^* = L^*A^*$ say, where $A^* = \Lambda A =$ an open set in M. The set $B = LA$ is open in L and $\Lambda B = (\Lambda L)(\Lambda A) = L^*A^* = A^*$. Thus the one-one correspondence induced by Λ between L and L^* pairs off sets open in L with sets open in L^* and is therefore a homeomorphism. It follows that L^* is closed.

2.4. *M^* is bicompact.* For let $\{G^*\}$ be a family of open sets covering M^*. Then the family of open sets $\{\Lambda^{-1}G^*\}$ covers M and a finite sub-family covering M can be chosen, say $\Lambda^{-1}G_1^*, \cdots, \Lambda^{-1}G_g^*$. The images of these sets under Λ gives a finite sub-family of $\{G^*\}$, namely $G_1^*, \cdots, G_g^*$, covering M^*. Hence M^* is bicompact. It can be proved just as readily that if the open set $A \subset M$ is an F_σ then ΛA is an F_σ.

2.5. A set $A \subset M$ will be called *invariant* if $A = A^1 (= A^2 = \cdots = A^{p-1})$. The sets $A, A^1, \cdots, A^{p-1}$ constitute the *A-sequence*. Let A be an open invariant set such that $L \subset A$, $M - A \neq 0$. A family $\{W_j^i\}$ $(i = 0, \cdots, p;\ j = 1, \cdots, w)$ of open sets will be called a *regular covering* of $M - A$ if (i) $M - A \subset \sum_{ij} W_j^i$ (ii) $\overline{W}_j^i L = 0$ (iii) each $\overline{W}_j^i$ meets at most one set in each $\overline{W}_k$-sequence $(k = 1, \cdots, w)$.

2.6. *There exist regular coverings of $M - A$.*

Proof. Let x be an arbitrary point of $M - A$. Using the fact that the points $x, x^1, \cdots, x^{p-1}$ are distinct and the fact that a bicompact space is normal, it can be seen from simple continuity considerations that there exists an open set V_x such that

$$(1) \qquad x \subset V_x, \qquad \bar{V}_x^i L = 0, \qquad \bar{V}^i \bar{V}^j = 0 \text{ if } i \neq j \qquad (i, j = 0, \cdots, p - 1).$$

Let a V_x be chosen for each $x \subset M - A$. Since $A = A^1 = \cdots = A^{p-1}$, each set $\sigma y (y \subset M)$ lies entirely in A or in $M - A$. Therefore $\Lambda(M - A) = \Lambda M - \Lambda A = M^* - A^*$, $(A^* = \Lambda A)$. Consequently the totality of open sets

ΛV_x covers the closed set $M^* - A^*$. Let a finite sub-covering be chosen, say $\Lambda V_1, \Lambda V_2, \cdots, \Lambda V_h$, (where $\Lambda V = \Lambda V_{x_i}$, $x_i \subset M - A$). Then $\Lambda^{-1}\Lambda V_1, \cdots, \Lambda^{-1}\Lambda V_h$ is a finite covering of $M - A$. Since $\Lambda^{-1}\Lambda V_i = \sigma V_i$, the sets V_j^i $(i = 0, \cdots, p - 1)$ $(j = 1, \cdots, h)$ cover $M - A$.

Consider the set V_1. Using again the fact that M is normal and that the sets of the $\bar{V}_2$-sequence (i.e. $\bar{V}_2, \bar{V}_2^1, \cdots, \bar{V}_2^{p-1}$) are mutually exclusive (by (1)), it is a simple exercise to show that V_1 is expressible as the sum of a finite number of open sets $X_1, \cdots, X_g$ such that each $\bar{X}$ meets at most one set in the $\bar{V}_2$-sequence. Each X can in turn be expressed as the sum of a finite number of open sets $Y_1, \cdots, Y_k$ such that each $\bar{Y}$ meets at most a single set of the $\bar{V}_3$-sequence; obviously each $\bar{Y}$ meets at most one set in the $\bar{V}_2$-sequence. Proceeding in this manner we finally obtain V_1 expressed as the sum of a finite number of open sets, say $U_1, \cdots, U_u$ such that each $\bar{U}_i$ meets at most one set in each V_j-sequence $(j = 2, \cdots, h)$. Hence if we replace the sets in the V_1-sequence by the totality of sets in the U_1-, U_2-, $\cdots$, U_u-sequences we obtain for $M - A$ a covering of the form

$$(2) \qquad \{U_j^i;\ V_2^i, V_3^i, \cdots, V_h^i\} \qquad (i = 0, \cdots, p - 1; j = 1, \cdots, u).$$

Each $\bar{U}_j$, and hence each $\bar{U}_j^i$, meets at most one set in each $\bar{V}_k$-sequence $(k > 2)$. Moreover each $\bar{U}_j$ (hence $\bar{U}_j^i$) meets at most one set in each $\bar{U}$-sequence since $\bar{U}_j \subset V_1$ and the sets $\bar{V}_1, \bar{V}_1^1, \cdots, \bar{V}_1^{p-1}$ are mutually exclusive. Moreover $\bar{U}_jL = 0$ since $\bar{V}_1L = 0$. If now we treat V_2 in the same manner relative to the covering (2) as we treated V_1 relative to the covering $\{V_j^i\}$ and if we continue in this manner step by step it is evident that the covering for $M - A$ which we obtain after h steps, is regular.

Let $\mathfrak{U}$ be a covering of M, that is a finite set of open sets or *vertices* whose sum is M. The covering whose vertices are the images under T^i of the vertices of $\mathfrak{U}$ will be denoted by $\mathfrak{U}^i$. $\mathfrak{U}$ is *invariant* if $\mathfrak{U}$ and $\mathfrak{U}^i$ are identical.

2.7. *Every covering $\mathfrak{U}$ of M possesses an invariant refinement.*

PROOF. Let $\mathfrak{B}$ be a common refinement of $\mathfrak{U}, \mathfrak{U}^1, \cdots, \mathfrak{U}^{p-1}$ and let $\mathfrak{W}$ consist of all the vertices of all the coverings $\mathfrak{B}, \mathfrak{B}^1, \cdots, \mathfrak{B}^{p-1}$. Obviously $\mathfrak{W} \subset \mathfrak{U}$, $\mathfrak{W}^1 = \mathfrak{W}$.

2.8. Let $\mathfrak{U}_L$ denote those vertices of an arbitrary covering $\mathfrak{U}$ which meet L, $\mathfrak{U}_N$ those which do not. The images under Λ of the vertices of $\mathfrak{U}$ form a covering of M^* which we shall denote by $\Lambda\mathfrak{U}$. Obviously the relation $\mathfrak{U} \subset \mathfrak{B}$ implies $\Lambda\mathfrak{U} \subset \Lambda\mathfrak{B}$.

2.9. Let $\mathfrak{U}$ be a covering of M such that (i) each vertex of $\mathfrak{U}_L$ is invariant; (ii) the vertices of $\mathfrak{U}_N$ fall into a finite number of iteration-sequences and each vertex of $\mathfrak{U}_N$ meets at most one vertex in each sequence; (iii) if $U \in \mathfrak{U}_N$, then $\bar{U}L = 0$. $\mathfrak{U}$ will be called a *regular covering of M*.

2.10. *Let* $\mathfrak{U} = \{U_1, \cdots, U_u\}$ *be a regular covering of* M. *A necessary and sufficient condition that* $\Lambda U_a = \Lambda U_b$ $(1 \leqq a, b \leqq u)$ *is that* U_b *be an image of* U_a *under a power of* T.

Proof. The sufficiency of the condition is obvious. Suppose that $\Lambda U_a = \Lambda U_b$. Then $\Lambda^{-1}\Lambda U_a = \Lambda^{-1}\Lambda U_b$, hence $\sigma U_a = \sigma U_b$. Therefore U_b must intersect some image of U_a, say $U_b U_a^i \neq 0$. Now suppose that $U_a^i \in \mathfrak{U}_N$. Then owing to the regularity of $\mathfrak{U}$ we have $U_a^i(U_b^1 + \cdots + U_b^{p-1}) = 0$. Hence $U_a^i(\sigma U_b) = U_a^i U_b \neq 0$. But since $U_a^i \subset \sigma U_a^i = \sigma U_a = \sigma U_b$, we have $U_a^i(\sigma U_b) = U_a^i$. Hence $U_a^i U_b = U_a^i$ and $U_a^i \subset U_b$. Since we are assuming that $U_a^i \in \mathfrak{U}_N$, it follows that $U_b \in \mathfrak{U}_N$ and therefore the argument can be repeated with U_a^i and U_b interchanged. Hence $U_b \subset U_a^i$, and $U_b = U_a^i$.—In case $U_a^i \in \mathfrak{U}_L$, we must have $U_b \in \mathfrak{U}_L$ and then $U_a = \sigma U_a = \sigma U_b = U_b$.

2.11. *Every invariant covering* $\mathfrak{U}$ *of* M *has a regular refinement.*

Proof. Suppose $\mathfrak{U}_L = \{U_1, \cdots, U_l\}$. Let $Q_i = U_i U_i^1 \cdots U_i^{p-1}$ $(i = 1, \cdots, l)$. The Q's are invariant open sets. Each Q meets L and $L \subset \sum_1^l Q_i = A$ (say). Let $\{V_j^i\}$ be a regular covering of $M - A$ (2.5). The V's together with the Q's form a covering $\mathfrak{V}$ of M. Let $\mathfrak{W} \subset \mathfrak{U}\mathfrak{V}$. We shall remove from $\mathfrak{W}$ each vertex which is contained in a Q, and replace the vertices which have been removed, by $Q_1, \cdots, Q_l$. The new covering thus formed consists of the Q's together with certain vertices, say $W_1, \cdots, W_t$ each one of which is contained in some V_j^i. Let all the images (under powers of T) of each vertex $W_1, \cdots, W_t$ be adjoined to $\mathfrak{W}$. The new covering $\mathfrak{W}_0$ thus formed consists of the Q's and the vertices W_j^i $(i = 0, \cdots, p - 1; j = 1, \cdots, t)$. It is clear that

$$(\mathfrak{W}_0)_L = \{Q_1, \cdots, Q_l\}, \qquad (\mathfrak{W}_0)_N = \{W_j^i\};$$

in fact each Q meets L, whereas each vertex W_j^i fails to meet L because it is contained in some V_k^h. Since $\{V_k^h\}$ is a regular covering of $M - A$, it follows that each vertex of $(\mathfrak{W}_0)_N$ (being contained in a V_k^h) meets at most one vertex of any sequence in $(\mathfrak{W}_0)_N$; moreover its closure fails to meet L. Therefore $\mathfrak{W}_0$ is regular. Since $Q_i \subset U_i$ $(i, 1, \cdots, l)$ and since, for $i = 1, \cdots, t$, $W_i \in \mathfrak{W} \subset \mathfrak{U}$ and $\mathfrak{U}$ is invariant, it follows that $\mathfrak{W}_0 \subset \mathfrak{U}$.

2.12. *Every covering of* M *possesses a regular refinement.* This follows from 2.7 and 2.11.

2.13. We now make the following assumptions concerning M:

I. Every open set in M is an F_σ.

II. dim $M \leqq n$.

I holds also for M^* (2.4). Consequently the results of dimension theory apply to M and M^*[2]. We shall show that dim $M^* \leqq n$. Let x^* be an arbitrary point of $M^* - L^*$, say $x^* = \Lambda x$, $x \subset M - L$. Let W be an open set containing x and such that $WL = 0$, $W^i W^j = 0$ when $i \neq j$ (cf. 2.6). Ob-

viously, if the domain of Λ is limited to W, Λ is one-one between W and ΛW. It follows readily that W and ΛW are homeomorphic (cf. 2.3) and therefore $\dim \Lambda W < n$. Since ΛW is a neighborhood of the arbitrary point $x^* \subset M^* - L^*$, we conclude that $\dim (M^* - L^*) \leqq n$. Since L and L^* are homeomorphic, $\dim L^* = \dim L \leqq n$. Since $M^* - L^*$ is an F_σ and L^* is closed, $\dim (M^* - L^*) + L^* = \dim M^* \leqq n$ ([6], p. 93 and [2]).

2.14. *Every covering* $\mathfrak{U}$ *possess a regular refinement of order* $\leqq n$.

PROOF. Let $\mathfrak{U}_0$ be refinement of $\mathfrak{U}$ such that those vertices of $\mathfrak{U}_0$ which meet a given vertex of $\mathfrak{U}_0$ are all contained in a single vertex of $\mathfrak{U}$. (Concerning the existence of such a refinement see [3], p. 628.) Let $\mathfrak{V}$ be a regular refinement of $\mathfrak{U}_0$. Obviously $\mathfrak{V}$ has the same property relative to $\mathfrak{U}$ as $\mathfrak{U}_0$ has. Suppose that

$$\mathfrak{V}_L = \{Q_1, \cdots, Q_q\}; \qquad \mathfrak{V}_N = \{V_j^i\} \qquad (i = 0, \cdots, p-1; j = 1, \cdots, v).$$

Since $\Lambda V_j^i = \Lambda V_j$, we have

$$\mathfrak{V}^* = \Lambda \mathfrak{V} = \{\Lambda Q_k ; \Lambda V_j\},\ \mathfrak{V}_N^* = \Lambda V_N = \{\Lambda V_j\} \ (k = 1, \cdots, q; j = 1, \cdots, v).$$

Since $\dim M^* \leqq n$, $\mathfrak{V}^*$ has a refinement $\mathfrak{W}^*$ of order $\leqq n$ ([6], p. 158; [2]). We may write

$$\mathfrak{W}^* = \{R_1^*, \cdots, R_r^*; W_1^*, \cdots, W_w^*\} = \{\Lambda R_1, \cdots, \Lambda R_r; \Lambda W_1, \cdots, \Lambda W_w\}$$

where the R_i and W_i are open sets of M and where the vertices of $\mathfrak{W}^*$ are so named that each R_i^* is contained in some Q_j, say $R_i^* \subset \Lambda Q_\alpha$ ($\alpha = \alpha(i)$) and each $W_i^* \subset \Lambda V_\beta$ ($\beta = \beta(i)$). We have $\Lambda^{-1} W_i^* \subset \Lambda^{-1}\Lambda B_\beta$, hence $\sigma W_i \subset \sigma V_\beta$. Let $X_i = (\sigma W_i) V_\beta$. Then $\Lambda X_i = (\Lambda \sigma W_i)(\Lambda V_\beta) = W_i^*(\Lambda V_\beta)$. Moreover

$$\sigma X_i = (\sigma\sigma W_i)(\sigma V_\beta) = (\sigma W_i)(\sigma V_\beta) = \sigma W_i .$$

Hence for each vertex W_i^* of $\mathfrak{W}^*$, there is an open set X_i such that (i) $\Lambda X_i = W_i^*$; (ii) $\sigma X_i = \sigma W_i$; (iii) $X_i \subset V_{\beta(i)}$. The open sets

$$\Lambda^{-1} R_1^*, \cdots, \Lambda^{-1} R_r^*; \qquad \Lambda^{-1} W_1^*, \cdots, \Lambda^{-1} W_w^*,$$

that is, by (ii), the sets

$$\sigma R_1, \cdots, \sigma R_r; \qquad \sigma X_1, \cdots, \sigma X_w$$

form a covering $\mathfrak{W}$ of M. For each i let the vertex σX_i of $\mathfrak{W}$ be now replaced by the p sets $X_i, X_i^1, \cdots, X_i^{p-1}$ and let the new covering thus formed be denoted by $\mathfrak{X}$. We assert that the order of $\mathfrak{X}$ is $\leqq n$. For consider the product

$$J = \sigma R_i \sigma R_j \cdots \sigma R_k X_q^a X_r^b \cdots X_s^c$$

of distinct vertices of $\mathfrak{X}$ and suppose that $J \neq 0$. Then the indices $q, r, \cdots, s$ must be distinct. For since X_q is contained in some vertex of the regular covering $\mathfrak{V}$ (by iii), the sets $X_q, X_q^1, \cdots, X_q^{p-1}$ are mutually exclusive and therefore if $q = r$, we must also have $a = b$ and $X_q^a = X_r^b$ contrary to hypothesis.

From this it follows (by i) that the images under Λ of the factors of J are *distinct* vertices of $\mathfrak{W}^*$ and since their product is $\subset \Lambda J \neq 0$, and $\mathfrak{W}^*$ is of order $\leqq n$, the factors of J must be $n + 1$ or fewer in number. Hence $\mathfrak{X}$ is of order $\leqq n$.

Let the invariant vertices $\sigma R_1, \cdots, \sigma R_r$ now be denoted by $S_1, \cdots, S_r$. Since each X_i is contained in some non-invariant vertex of $\mathfrak{V}_N$, we have $X_j^i L = 0$. Consequently $L \subset S_1 + \cdots + S_r$. It is not necessarily true, however, that each S meets L, and therefore $\mathfrak{W}$ need not be regular.

Let us recall that $\Lambda R_i \subset Q_{\alpha(i)}$ and that Q_α is invariant. Then

$$S_i = \sigma R_i = \Lambda^{-1}\Lambda R_i = \Lambda^{-1} R_i^* \subset \Lambda^{-1} Q_\alpha = Q_\alpha$$

so that each S is contained in some Q. We may assume that the S's are so named that $S_1, \cdots, S_l$ meet L, $S_{l+1}, \cdots, S_r$ do not. Assume also that

$$S_i \subset Q_{a_i} \qquad (i = 1, \cdots, l), \qquad S_{l+i} \subset Q_{b_i} \qquad (i = 1, \cdots, r - l).$$

We assert that $Q_{b_j}(\sum_{i=1}^{l} Q_{a_i}) \neq 0$ $(j = 1, \cdots, r - l)$. For otherwise LQ_{b_j} is covered by S's from the set $S_{l+1}, \cdots, S_r$ whereas these S's do not meet L. Suppose then that Q_{b_j} meets Q_{a_k} $(k = k(j), j = 1, \cdots, r - l; 1 \leqq k \leqq l)$. Let the S's in Q_{b_1} having subscripts $> l$ be grouped with $S_{k(1)}$, those in Q_{b_2} but not in Q_{b_1}, with $S_{k(2)}$ and so on. In this way the S's can be arranged in l families, the i^{th} family consisting of S_i together with certain S's which are contained in Q's which meet Q_{a_i}. Let Y_i be the sum of the S's in the i^{th} family and let the vertices $S_1, \cdots, S_r$ of $\mathfrak{X}$ be replaced by the open sets $Y_1, \cdots, Y_l$. We obtain a new covering

$$\mathfrak{X}_0 = \{Y_1, \cdots, Y_l; X_j^i\} \qquad (i = 0, \cdots, p - 1; j = 1, \cdots, w).$$

$\mathfrak{X}_0$ is regular. For each Y_i is invariant and meets L, and the X_j^i's derive their required properties from the fact that each is contained in a non-invariant vertex of the regular $\mathfrak{V}$. Furthermore, $\mathfrak{X}_0 \subset \mathfrak{U}$. To prove this it is only necessary to examine the Y's since $X_j^i \subset V_{\beta(j)}^i \in \mathfrak{V} \subset \mathfrak{U}$. But Y_i is contained in the sum of certain Q's which meet Q_{a_i} and these Q's together with Q_{a_i} are contained in a single vertex of $\mathfrak{U}$, on account of the way $\mathfrak{U}_0$ was chosen and the fact that $Q_i \in \mathfrak{V} \subset \mathfrak{U}_0$ $(i = 1, \cdots, l)$. Finally, since $\mathfrak{X}$ is of order $\leqq n$, and since $\mathfrak{X}_0$ was obtained from $\mathfrak{X}$ by the operation (applied perhaps several times) of replacing a set of vertices of their sum, $\mathfrak{X}_0$ is also of order $\leqq n$. Therefore $\mathfrak{X}_0$ is the required refinement of $\mathfrak{U}$.

2.15. Let a be a point of L and A a neighborhood of a (i.e. an open set containing a). There exists an invariant neighborhood B of a, such that $B \subset A$. In fact, $AA^1 \cdots A^{p-1}$ is such a neighborhood.

2.16. We shall say that a covering $\mathfrak{U}$ of M is *special* if (i) $\mathfrak{U}$ is regular; (ii) every $\mathfrak{U}$-simplex with vertices in $\mathfrak{U}_L$ is in L;[5] (iii) every $\mathfrak{U}$-simplex not in L is of dimension $\leqq n$.

[5] The simplex $(U_0 U_1 \cdots U_h)$ is in L if $U_0 U_1 \cdots U_h L \neq 0$.

A special $\mathfrak{U}$ has the property that the totality of $\mathfrak{U}$-simplexes which are in L is identical with the totality of $\mathfrak{U}$-simplexes with vertices in $\mathfrak{U}_L$. This follows immediately from (ii) and the fact that no vertex of $\mathfrak{U}_N$ meets L.

2.17. *Every covering of M has a special refinement.*

PROOF. Let $\mathfrak{U}$ be a covering of M and $\mathfrak{B}$ a refinement of $\mathfrak{U}$ such that for an arbitrary $\mathfrak{B}$-simplex E, there exists a vertex of $\mathfrak{U}$ containing all the vertices of E.[6] Choose a refinement $\mathfrak{W}$ of $\mathfrak{U}$ which is regular and of order $\leqq n$ (2.14). Let $\mathfrak{W}_L = \{O_1, \cdots, O_l\}$, and let x_i be a point in W_i $(i = 1, \cdots, l)$. Let X_i be an invariant neighborhood of x_i such that $\bar{X}_i \subset O_i$ and $X_i W_N = 0$ where W_N means the sum of the vertices of $\mathfrak{W}_N$. That the X's exist follows readily from 2.5 and the definition of regular covering. Let $O_{i_1}, O_{i_2}, \cdots, O_{i_a}$ $(a = a(i))$ be the O's which meet O_i. Let $Z_i = X_{i_1} + X_{i_2} + \cdots + X_{i_a}$ and let $O_i' = O_i + Z_i$ $(i = 1, \cdots, l)$. We shall show that the covering $\mathfrak{W}'$ obtained from $\mathfrak{W}$ by replacing each O_i by O_i' is special. Obviously $\mathfrak{W}'$ is regular and $\mathfrak{W}_L' = \{O_i'\}$, $W_N' = W_N =$ (say) $\{W_1, \cdots, W_m\}$. Suppose $O_a' O_b' \cdots O_c' \neq 0$. Then $Z_a \supset X_b + \cdots + X_c$ so that

$$O_a' \supset X_a + Z_a \supset X_a + X_b + \cdots + X_c$$

and similarly for $O_b', \cdots, O_c'$. Hence

$$O_a' O_b' \cdots O_c' \supset X_a + X_b + \cdots + X_c$$

and therefore $O_a' O_b' \cdots O_c' L \neq 0$. Consequently (ii) is satisfied. Next let E be a $\mathfrak{W}'$-simplex not in L, say $E = (O_a', \cdots, O_c', W_p, \cdots, W_q)$. At least one W must enter the symbol for E, otherwise E would be in L. Let

$$(1) \qquad J = O_a' \cdots O_c' W_p \cdots W_q = (O_a + Z_a) \cdots (O_c + Z_c) W_p \cdots W_q.$$

Since no W meets any X and therefore, none meets any Z, it follows that if J be expanded as a sum, all terms of the form $O_s \cdots O_t Z_i \cdots Z_j W_p \cdots W_q$ are zero. Hence

$$(2) \qquad J = O_a \cdots O_c W_p \cdots W_q \neq 0.$$

Obviously the factors in (2) are distinct because those in (1) are. Since $\mathfrak{W}$ is of order $\leqq n$, there can be at most $n + 1$ factors in (1) and (2) and therefore E is of dimension $\leqq n$, and (iii) is satisfied. It remains to be proved that $\mathfrak{W}'$ is a refinement of $\mathfrak{U}$. For each O_i' there exists a $\mathfrak{W}$-simplex the sum of whose vertices contains O_i'. Hence, by the choice of $\mathfrak{B}$ and $\mathfrak{W}$, O' is contained in a vertex of $\mathfrak{U}$ and hence $\mathfrak{W}' \subset \mathfrak{U}$.

2.18. Unless it is stated to the contrary, let it be understood from now on that "covering of M" means a *special* covering of M. Since these coverings form a complete family S of M (by 2.17), the homology properties of M can be studied solely in terms of the coverings in S. Moreover, it is easy to see

[6] Cf. [3] page 628.

that the totality S^* of coverings of M^* of the form $\Lambda\mathfrak{U}(\mathfrak{U} \epsilon S)$ is a complete family for M^* and consequently the homology properties of M^* can be studied in terms of the coverings in S^*.

2.19. *Let $\mathfrak{U}, \mathfrak{V}$ be coverings of M, $\mathfrak{U} \subset \mathfrak{V}$. There exists an "invariant projection" of $\mathfrak{U}$ into $\mathfrak{V}$, that is, a projection π such that $\pi U^i = (\pi U)^i$ $(i = 0, \cdots, p-1)$ for every $\mathfrak{U}^i \epsilon U$.*

PROOF. Consider the arbitrary vertex $O \epsilon \mathfrak{U}_L$. There must exist a vertex $Q \epsilon \mathfrak{V}_L$ such that $O \subset Q$. Let $\pi O = Q$. Since O and Q are invariant we have $\pi O^i = (\pi O)^i$ $(i = 0, \cdots, p-1)$. Suppose $\mathfrak{U}_N = \{U_j^i\}$ $(i = 0, \cdots, p-1;$ $j = 1, \cdots, u)$. There exists for each U_j^i a vertex $V_{a(j)}$ in $\mathfrak{V}_L$ or $\mathfrak{V}_N$ such that $U_j \subset V_{a(j)}$. Then $U_j^i \subset V_{a(j)}^i$ and if we put $\pi U_j^i = V_{a(j)}^i$, we have $\pi(U_j^i)^k = (\pi U_j^i)^k$ $(k = 0, \cdots, p-1)$ and hence π is invariant.

2.20. Let it be understood from now on unless stated to the contrary, that all projections of (special) coverings of M are invariant.

Let $\mathfrak{V}$ be a refinement of $\mathfrak{U}$ and let $\pi = \pi(\mathfrak{V}, \mathfrak{U})$. In a natural manner we can associate to π a projection π_λ of $\Lambda\mathfrak{V}$ into $\Lambda\mathfrak{U}$ by means of the formula $\pi_\lambda\Lambda\mathfrak{V} = \Lambda\pi\mathfrak{V}$. We must show that this definition of π_λ is self-consistent; that is, if V_1, V_2 are two vertices of $\mathfrak{U}$ such that $\Lambda V_1 = \Lambda V_2$, then we must show that $\pi_\lambda\Lambda V_1 = \pi_\lambda\Lambda V_2$. But the relation $\Lambda V_1 = \Lambda V_2$ implies that $V_2 = V_1^a$ (2.10); then since π is invariant we have $\pi V_2 = (\pi V_1)^a$. Hence $\Lambda\pi V_2 = \Lambda(\pi V_1)^a = \Lambda\pi V_1$ so that $\pi_\lambda V_2 = \pi_\lambda V_1$.

2.21. *Let $\mathfrak{U}, \mathfrak{V}$ be coverings of M such that $\Lambda\mathfrak{V} \subset \Lambda\mathfrak{U}$. There exists a covering $\mathfrak{V}_0$ such that $\Lambda\mathfrak{V}_0 = \Lambda\mathfrak{V}$, $\mathfrak{V}_0 \subset \mathfrak{U}$.*

PROOF. Let

$$\mathfrak{U}_L = \{O_k\}, \quad \mathfrak{U}_N = \{U_j^i\}; \quad \mathfrak{V}_L = \{Q_q\}, \quad \mathfrak{V}_N = \{V_h^i\} \quad (i = 0, \cdots, p-1).$$

Since $\mathfrak{U}$ and $\mathfrak{V}$ are regular, each ΛQ must be contained in a ΛO, say $\Lambda Q_q \subset \Lambda O_{a(q)}$. Let $\Lambda V_1, \cdots, \Lambda V_{h_0}$ be the ΛV_h's which are contained in ΛO's, say $\Lambda V_h \subset \Lambda O_{b(h)}$ $(h = 1, \cdots, h_0)$. Each remaining ΛV_h is contained in a ΛU, say $\Lambda V_h \subset \Lambda U_{c(h)}$ $(h = h_0 + 1, \cdots, l)$. Since Λ induces a homeomorphism (cf. 2.13) between U_c and ΛU_c, we may choose an open set V_{g0} such that $V_{g0} \subset U_{c(g)}$, $\Lambda V_{g0} = \Lambda V_g$ $(g = h_0 + 1, \cdots, l)$. Let

$$\mathfrak{V}_0 = \{Q_k\,;\, \sigma V_h\,;\, V_{g0}^i\}$$
$$(h = 1, \cdots, h_0\,;\, g = h_0 + 1, \cdots l;\, i = 0, \cdots, p-1).$$

$\mathfrak{V}_0$ is a covering of M and it is readily seen that $\Lambda V_0 = \Lambda V$. Moreover

$$Q_q = \sigma Q_q = \Lambda^{-1}\Lambda Q_q \subset \Lambda^{-1}\Lambda O_{a(q)} = \sigma O_{a(q)} = O_{a(q)}\,;$$
$$\sigma V_h = \Lambda^{-1}\Lambda V_h \subset \Lambda^{-1}\Lambda O_{b(h)} = \sigma Q_{b(h)} = Q_{b(h)}\,;\, V_{g0}^i \subset V_{c(g)}^i\,.$$

Hence $\mathfrak{V}_0 \subset \mathfrak{U}$.

3. Homologies in M and M^*

3.1. Let $\mathfrak{U}$ be a covering of M. We shall describe certain relations between $\mathfrak{U}$-chains of M and the $\Lambda\mathfrak{U}$-chains of M^*.

Let $E = (U_0, U_1, \cdots, U_h)$ be an $(h, \mathfrak{U})$-simplex. Then $(U_0^i, U_1^i, \cdots, U_h^i)$ is an $(h, \mathfrak{U})$-simplex and will be denoted by E^i. Moreover, the vertices $\Lambda\mathfrak{U}_0, \cdots, \Lambda\mathfrak{U}_h$ of $\Lambda\mathfrak{U}$ are distinct. For since $U_aU_b \neq 0$ $(0 \leqq a, b \leqq h)$, U_b can not be of the form U_a^i if $U_b \in \mathfrak{U}_N$; nor can $U_b = U_a^i$ if U_a is invariant for then we would have $U_b = U_a$. Consequently (2.10) the sets $\Lambda\mathfrak{U}_0, \cdots, \Lambda\mathfrak{U}_h$ are distinct; moreover their product is ΛJ where $J = \prod_{i=0}^h U_i \neq 0$. Hence $(\Lambda\mathfrak{U}_0, \cdots, \Lambda\mathfrak{U}_h)$ is an $(h, \Lambda\mathfrak{U})$-simplex and will be denoted by ΛE. We thus extend the domain of operation of Λ to include the simplexes of an arbitrary (special) covering of M.

3.2. *Let* $\Lambda\mathfrak{U} = \mathfrak{U}^*$ *and let* E^* *be an* $(h, \mathfrak{U}^*)$*-simplex. There exists at least one* $(h, \mathfrak{U})$*-simplex* E *such that* $\Lambda E = E^*$.

Proof. Let $E^* = (U_0^*, \cdots, U_h^*)$, $U_i^* = \Lambda U_i$, $U_i \in \mathfrak{U}$. There exist integers $a_0, \cdots, a_h$ such that $\prod_{j=0}^h U_j^{a_j} \neq 0$. For since $\prod_{i=0}^h U_i^* \neq 0$, we have $\prod_{j=0}^h (\Lambda^{-1}U_j^*) = \prod_{j=0}^h (\sigma U_j) \neq 0$. But

$$\prod_{j=0}^h (\sigma U_j) = \sum_{i_0, \cdots, i_h = 0}^{p-1} U_0^{i_0} U_1^{i_1} \cdots U_h^{i_h};$$

therefore at least one term in the sum, say $U_0^{a_0} \cdots U_h^{a_h}$ is $\neq 0$. Since $\Lambda U_i^{a_i} = \Lambda U_i = U_i^*$, the vertices $U_0^{a_0}, \cdots, U_h^{a_h}$ are distinct (2.10) and therefore determine an $(h, \mathfrak{U})$ simplex E. Obviously $\Lambda E = E^*$.

3.3. *Let* E_1, E_2 *be* $(h, \mathfrak{U})$*-simplexes. A necessary and sufficient condition that* $\Lambda E_1 = \Lambda E_2$ *is that for some* a $(0 \leqq a \leqq p - 1)$, $E_1 = E_2^a$.

Proof. The sufficiency is obvious. Assume then that $\Lambda E_1 = \Lambda E_2$. Suppose that $E_1 = (O_{11}, \cdots, O_{1p}, U_{11}, \cdots, U_{1q})$ where $p + q = h + 1$ and $O_{1i} \in \mathfrak{U}_L$, $U_{1i} \in \mathfrak{U}_N$. It is clear then that E_2 must be of the form

$$(O_{21}, \cdots, O_{2p}, U_{21}, \cdots, U_{2q})$$

where $\Lambda O_{1i} = \Lambda O_{2i}$ $(i = 1, \cdots, p)$, $\Lambda U_{1i} = \Lambda U_{2i}$ $(i = 1, \cdots, q)$. Since the O's are invariant, we have immediately $O_{1i} = O_{2i}$ $(i = 1, \cdots, q)$. As for the U's, we have

$$\sigma U_{1j} = \Lambda^{-1}\Lambda U_{1j} = \Lambda^{-1}\Lambda U_{2j} = \sigma U_{2j}.$$

Hence $\Lambda U_{1j} = \Lambda\sigma U_{1j} = \Lambda\sigma U_{2j} = \Lambda U_{2j}$ and it follows from 2.10, that $U_{1j} = U_{2j}^{a_j}$ $(0 \leqq a_j \leqq p - 1)$. Hence, since $U_{11}U_{12} \neq 0$, we have $U_{21}^{a_1}U_{22}^{a_2} \neq 0$ and therefore $U_{21}U_{22}^{a_2 - a_1} \neq 0$. Suppose $a_2 - a_1 \neq 0 \bmod p$. Then since $U_{21}U_{22} \neq 0$, U_{21} is intersected by two vertices of the sequence $U_{22}^0, \cdots, U_{22}^{p-1}$ which is impossible because $\mathfrak{U}$ is regular. Therefore $a_1 = a_2 \pmod p$ and in fact, we obtain, on repeating the argument, $a_1 = a_2 = \cdots = a_q = a$ (say), mod p.

Therefore $U_{1j} = U^a_{2j}$. Since also $O_{1i} = O^a_{2i}$, we have $E_1 = E^a_2$. We may of course reduce a modulo p.

3.4. Let E be a $\mathfrak{U}$-simplex not in L. *The simplexes* $E, E^1, \cdots, E^{p-1}$ *are distinct.* For suppose $E = (U_0, \cdots, U_h)$ and suppose, as we may, that U_0 is in $\mathfrak{U}_N$. If $E = E^i$, then U_0 is a vertex of E^i; hence U_0 meets each vertex of E^i; in particular U_0 meets U_0^i, and this is impossible since $\mathfrak{U}$ is special.

3.5. Let $C(\mathfrak{U}) = \sum x_j E_j$ be an $(h, \mathfrak{U})$-chain with coefficients x_j in a field $\mathfrak{f}$. We shall denote the $(h, \mathfrak{U})$-chain $\sum x_j E_j^i$ by $C^i(\mathfrak{U})$ and the $(h, \Lambda\mathfrak{U})$-chain $\sum x_j \Lambda E_j$ by $\Lambda C(\mathfrak{U})$. Obviously $\Lambda C(\mathfrak{U}) = \Lambda C^i(\mathfrak{U})$. Let the chain $\sum_{i=0}^{p-1} C^i(\mathfrak{U})$ be denoted by $\sigma C(\mathfrak{U})$.

Let $E = (U_0, \cdots, U_h)$ be an $(h, \mathfrak{U})$-simplex in L. The vertices of E are in $\mathfrak{U}_L$ and each vertex is therefore invariant. Hence E itself is invariant—i.e. $E = E^i$ $(i = 1, 2, \cdots)$. Conversely, if $E = E^i$, it follows from 3.4 that $E \subset L$. We conclude that a chain $C(\mathfrak{U})$ will be simplex-wise invariant if and only if $C(\mathfrak{U}) \subset L$.

3.6. A further remark concerning the chains in L. It is clear from the properties of special coverings that the totality of $\mathfrak{U}^*$-simplexes in L^* ($\mathfrak{U}^* = \Lambda\mathfrak{U}$) is identical with the image under Λ of the totality of $\mathfrak{U}$-simplexes in L. Moreover, the vertices of $\Lambda\mathfrak{U}_L$ have precisely the same intersection relations among themselves as the corresponding vertices in $\mathfrak{U}_L$. Therefore, Λ induces a one-one correspondence between the $\mathfrak{U}$-chains in L and the $\mathfrak{U}^*$-chains in L^*, of such a nature that under Λ all homology relations are preserved.

3.7. Let $E = (U_0, \cdots, U_h)$ be an $(h, \mathfrak{U})$-simplex. Then

$$\phi E = \sum (-1)^k (U_0, \cdots, U_{k-1}, U_{k+1}, \cdots, U_h)$$

and

$$\Lambda\phi E = \sum (-1)^k (\Lambda U_0, \cdots, \Lambda U_{k-1}, \Lambda U_{k+1}, \cdots, \Lambda U_h)$$

so that

$$\Lambda\phi C(\mathfrak{U}) = \phi\Lambda C(\mathfrak{U}) \tag{1}$$

for every $C(\mathfrak{U})$. Moreover, if $C(\mathfrak{U}) \subset A$, then $\Lambda C(U) \subset \Lambda A$. Therefore, if $D(\mathfrak{U})$ is a cycle mod B in $A(B \subset A \subset M)$, then $\Lambda D(\mathfrak{U})$ is a cycle mod ΛB in ΛA. If $D(\mathfrak{U}) \sim 0$ mod B in A, then $\Lambda D(\mathfrak{U}) \sim 0$ mod ΛB in ΛA.

Let $\mathfrak{U}^* = \Lambda\mathfrak{U}$ and let E^* be an $(h, \mathfrak{U}^*)$-simplex. By 3.2 there exists an $(h, \mathfrak{U})$-simplex E such that $\Lambda E = E^*$. Let σE be denoted by $\Lambda^{-1}E^*$. If E_1 is a second $(h, \mathfrak{U})$-simplex such that $\Lambda E_1 = E^*$, then $E_1 = E^a$ by 3.3, and therefore $\sigma E_1 = \sigma E$. Consequently the chain $\Lambda^{-1}E^*$ is uniquely determined. If $C^*(\mathfrak{U}^*) = \sum x_j E_j^*$ is an $(h, \mathfrak{U}^*)$-chain, we shall denote $\sum x_j \Lambda^{-1} E_j^*$ by $\Lambda^{-1}C^*(\mathfrak{U}^*)$. The relations

$$\Lambda^{-1}\Lambda C(\mathfrak{U}) = \sigma C(\mathfrak{U}) \tag{2}$$

$$\Lambda\Lambda^{-1}D^*(\mathfrak{U}^*) = pD^*(\mathfrak{U}^*) \tag{3}$$

hold for every $C(\mathfrak{U})$ and $D^*(\mathfrak{U}^*)$.

3.8. Again let E^* be an $(h, \mathfrak{U}^*)$-simplex, and choose E such that $\Lambda E = E^*$. Using the relations (1) and (2) of 3.7 and the obvious relation $\phi E^i = (\phi E)^i$, we have

$$\Lambda^{-1}\phi E^* = \Lambda^{-1}\phi\Lambda E = \Lambda^{-1}\Lambda\phi E = \sigma\phi E = \phi\sigma E = \phi\Lambda^{-1}E^*$$

and from this follows the general relation

$$\Lambda^{-1}\phi D^*(\mathfrak{U}^*) = \phi\Lambda^{-1}D^*(\mathfrak{U}^*)$$

where $D^*(\mathfrak{U}^*)$ is an arbitrary $\mathfrak{U}^*$-chain. If $D^*(\mathfrak{U}^*) \subset A^*$, $A^* = \Lambda A$, then $\Lambda^{-1}D^*(\mathfrak{U}^*) \subset \sigma A$. Therefore, if $D^*(\mathfrak{U}^*)$ is a cycle mod B^* in A^* ($B^* = \Lambda B \subset A^* = \Lambda A$), then $\Lambda^{-1}D^*(\mathfrak{U}^*)$ is a cycle mod σB in σA, and if $D^*(\mathfrak{U}^*) \sim 0$ mod B^* in A^*, then $\Lambda^{-1}D^*(\mathfrak{U}^*) \sim 0$ mod σB in σA.

3.9. We have seen (3.4) that if an $(h, \mathfrak{U})$-simplex E is not in L, the simplexes $E, E^1, \cdots, E^{p-1}$ are distinct; obviously none of these simplexes are in L. It is clear then that the $(h, \mathfrak{U})$-simplexes which are not in L can be represented without repetition by E_j^i ($i = 0, \cdots, p-1$; $j = 1, \cdots, \alpha_h$).

Let $X(\mathfrak{U})$ be an arbitrary $(h, \mathfrak{U})$-chain. We shall denote the chain $X(\mathfrak{U}) - X^1(\mathfrak{U})$ by $\bar{\sigma}(\mathfrak{U})$. We shall now use the symbol ρ to mean either σ or $\bar{\sigma}$ and shall agree that if $\rho = \sigma$ (or $\bar{\sigma}$) then $\bar{\rho} = \bar{\sigma}$ (or σ). It is readily seen that in any case

$$\rho\bar{\rho}X(\mathfrak{U}) = \bar{\rho}\rho X(\mathfrak{U}) = 0. \tag{1}$$

A chain of the form $\rho X(\mathfrak{U})$ will be called a $\mathfrak{U}$-chain of type ρ.

3.10. *Let $Y(\mathfrak{U})$ be an $(h, \mathfrak{U})$-chain containing no simplex in L. A necessary and sufficient condition that $Y(\mathfrak{U})$ be of type ρ is that $\bar{\rho}Y(\mathfrak{U}) = 0$.*

Proof. The necessity follows from (1), 3.9. Suppose then that $\bar{\rho}Y(\mathfrak{U}) = 0$. In any case we may write (3.9)

$$Y(\mathfrak{U}) = \sum_{j=1}^{\alpha_h}\sum_{i=0}^{p-1} y_j^i E_j^i.$$

The condition $\bar{\rho}Y(\mathfrak{U})$ then becomes

$$\sum_{i,j} y_j^i \bar{\rho}E_j^i = 0. \tag{1a}$$

We shall consider separately the cases $\bar{\rho} = \bar{\sigma}$ and $\bar{\rho} = \sigma$.

Suppose $\bar{\rho} = \bar{\sigma}$. Then (1a) becomes

$$\sum_{i,j} y_j^i(E_j^i - E_j^{i+1}) = 0$$

where the upper indices are to be reduced mod p. Hence

$$\sum_{i,j} (y_j^i - y_j^{i-1}) E_j^i = 0$$

and hence $y_j^i = y_j^{i-1}$ $(i = 0, \cdots, p-1)$. Therefore $y_j^0 = y_j^1 = \cdots = y_j^{p-1} = y_j$ (say), so that

$$Y(\mathfrak{U}) = \sigma X(\mathfrak{U})$$

where $X(\mathfrak{U}) = \sum_{j=1}^{\alpha} y_j E_j$, which concludes the proof for the case $\bar{\rho} = \bar{\sigma}$.

Suppose $\bar{\rho} = \sigma$. Then (1a) becomes

$$\sum_{i,j} y_j^i \sigma E_j^i = \sum_{i,j} y_j^i \sigma E_j = \sum_{j=1}^{\alpha} \left(\sum_{i=0}^{p-1} y_j^i \right) \sigma E_j = 0.$$

Hence

$$\sum_{i=0}^{p-1} y_j^i = 0 \tag{2}$$

Let

$$X(\mathfrak{U}) = \sum_{j=1}^{\alpha} \sum_{k=0}^{p-1} \sum_{i=0}^{k} y_j^i E_j^k.$$

Then

$$\begin{aligned}
X(\mathfrak{U}) - X^1(\mathfrak{U}) &= \sum_{j=1}^{\alpha} \sum_{k=0}^{p-1} \sum_{i=0}^{k} y_j^i (E_j^k - E_j^{k+1}) \\
&= \sum_{j=1}^{\alpha} \left[\sum_{k=0}^{p-1} \sum_{i=0}^{k} y_j^i E_j^k - \sum_{k=1}^{p} \sum_{i=0}^{k-1} y_j^i E_j^k \right] \\
&= \sum_{j=1}^{\alpha} \left[y_j^0 E_j^0 + \sum_{k=1}^{p-1} \left(\sum_{i=0}^{k} y_j^i - \sum_{i=0}^{k-1} y_j^i \right) E_j^k - \sum_{i=0}^{p-1} y_j^i E_j^p \right] \\
&= \sum_{j=1}^{\alpha} \left[\sum_{k=0}^{p-1} y_j^k E_j^k - \sum_{i=0}^{p-1} y_j^i E_j^p \right] \\
&= \sum_{j=1}^{\alpha} \sum_{k=0}^{p-1} y_j^k E_j^k \qquad \text{(on account of 2)} \\
&= Y(\mathfrak{U}).
\end{aligned}$$

Hence $Y(\mathfrak{U}) = \bar{\sigma} X(\mathfrak{U})$ and the proof is complete.

3.11. Let B be an invariant set, possibly null. We shall say that the chain $Y(\mathfrak{U})$ is of type (ρ, B) if there exists a chain $X(\mathfrak{U})$ such that $Y(\mathfrak{U}) = \rho X(\mathfrak{U})$ mod B. A chain of type $(\rho, 0)$ is obviously of type ρ. *If $L \subset B$, a necessary and sufficient condition that $Y(\mathfrak{U})$ be of type (ρ, B) is that $\bar{\rho} Y(\mathfrak{U}) = 0$ mod B.*

Proof. Suppose $Y(\mathfrak{U}) = \rho X(\mathfrak{U})$ mod B. Then since B is invariant, we have $Y^i(\mathfrak{U}) = \rho X^i(\mathfrak{U})$ mod B. Hence $\bar{\rho} Y(\mathfrak{U}) = \bar{\rho}\rho X(\mathfrak{U}) = 0$ mod B (3.9).

Conversely, suppose $\bar{\rho} Y(\mathfrak{U}) = 0$ mod B. Since $L \subset B$, we may write $Y(\mathfrak{U}) =$

$'Y(\mathfrak{U})$ mod B where $'Y(\mathfrak{U})$ is a chain containing no simplexes in B. We have also $Y^i(\mathfrak{U}) = {'Y^i}(\mathfrak{U})$ mod B (since $B^i = B$), hence $\bar{\rho}'Y(\mathfrak{U}) = \bar{\rho}Y(\mathfrak{U}) = 0$ mod B. Hence by 3.10, $'Y(\mathfrak{U})$ is of the form $\rho X(\mathfrak{U})$ and we have $Y(\mathfrak{U}) = \rho X(\mathfrak{U})$ mod B.

3.12. Let B be an invariant set, possibly null, and let $\Gamma_h(\mathfrak{U})$ be a cycle mod B in A ($B \subset A \subset M$) of type (ρ, B). If there exists in A an $(h + 1, \mathfrak{U})$-chain $H(\mathfrak{U})$ of the same type as $\Gamma_h(\mathfrak{U})$—that is, of type (ρ, B), such that $\phi H(\mathfrak{U}) = \Gamma(\mathfrak{U})$ mod B, we shall write $\Gamma_h(\mathfrak{U}) \simeq 0$ mod B in A. If $A = M$ or if $\Gamma_h(\mathfrak{U}) = 0$ mod B, we shall write simply $\Gamma_h(\mathfrak{U}) \simeq 0$ mod B. It is quite easy to see that homologies of the sort we have just defined obey the same formal laws as ordinary homologies.

3.13. Let A, B be invariant sets such that $B \subset A \subset M$ and let $\Gamma_h(\mathfrak{U})$ be a cycle mod B in A of type (ρ, B). Suppose there exists in A a chain $X_{h+1}(\mathfrak{U})$ such that $\phi X_{h+1}(\mathfrak{U}) = \Gamma_h(\mathfrak{U})$ mod B. Then $\bar{\rho}X_{h+1}(\mathfrak{U})$ is a cycle mod B in A of type $\bar{\rho}$ (hence of type $(\bar{\rho}, B)$. For, we have (say) $\Gamma_h(\mathfrak{U}) = \rho Y(\mathfrak{U})$ mod B. Hence

$$\phi\bar{\rho}X_{h+1}(\mathfrak{U}) = \bar{\rho}\phi X_{h+1}(\mathfrak{U}) = \bar{\rho}\rho Y(\mathfrak{U}) = 0 \text{ mod } B.$$

Since $A^i = A$, we have $X^i(\mathfrak{U}) \subset A$, hence $\bar{\rho}X_{h+1}(\mathfrak{U}) \subset A$.

3.14. Let B be an invariant set containing L and let $\Gamma_h(\mathfrak{U})$ be a cycle mod B of type (ρ, B), say $\Gamma_h(\mathfrak{U}) = \rho X_h(\mathfrak{U})$ mod B. Let $\Gamma_{h-1}(\mathfrak{U}) = \phi X_h(\mathfrak{U})$. The absolute cycle $\Gamma_{h-1}(\mathfrak{U})$ is of type $(\bar{\rho}, B)$. For

$$\rho\Gamma_{h-1}(\mathfrak{U}) = \rho\phi X = \phi\rho X = \phi\Gamma_h(\mathfrak{U}) = 0 \text{ mod } B.$$

It will be convenient to indicate the relation which exists between $\Gamma_h(\mathfrak{U})$ and $\Gamma_{h-1}(\mathfrak{U})$ by writing $\Gamma_h(\mathfrak{U})\colon \Gamma_{h-1}(\mathfrak{U})$. Obviously $\Gamma_{h-1}(\mathfrak{U})$ is not uniquely determined since $X_h(\mathfrak{U})$ can in general be chosen in a multiplicity of ways.

3.15. Let A, B be sets such that $L \subset B = B^1 \subset A \subset M$. Let $\Gamma_h(\mathfrak{U})$ be a cycle mod B in A of type (ρ, B) and let $\Gamma_{h-1}(\mathfrak{U})$ be a cycle in A of type $(\bar{\rho}, B)$ such that $\Gamma_h(\mathfrak{U})\colon \Gamma_{h-1}(\mathfrak{U})$. If $\Gamma_h(\mathfrak{U}) \simeq 0$ mod B in A, then $\Gamma_{h-1}(\mathfrak{U}) \simeq 0$ mod B in A.

PROOF. If $\Gamma_h(\mathfrak{U}) \simeq 0$ mod B in A, there exists a chain $K_{h+1}(\mathfrak{U})$ in A such that $\rho K_{h+1}(\mathfrak{U}) \subset A$ and $\phi\rho K_{h+1}(\mathfrak{U}) = \Gamma_h(\mathfrak{U})$ mod B. In particular if $\Gamma_h(\mathfrak{U}) = 0$ mod B (see 3.11) we shall take $K_{h+1}(\mathfrak{U}) = 0$. In any case

$$\phi\rho K_{h+1}(\mathfrak{U}) = \Gamma_h(\mathfrak{U}) = \rho X_h(\mathfrak{U}) \text{ mod } B.$$

Let $H_h(\mathfrak{U}) = X_h(\mathfrak{U}) - \phi K_{h+1}(\mathfrak{U})$. Then $H_h(\mathfrak{U}) \subset A$ and

$$\phi H_h(\mathfrak{U}) = \phi X_h(\mathfrak{U}) = \Gamma_{h-1}(\mathfrak{U}) \text{ mod } B.$$

Moreover

$$\rho H_h(\mathfrak{U}) = \rho X_h(\mathfrak{U}) - \phi\rho K_{h+1}(\mathfrak{U}) = 0 \text{ mod } B.$$

Hence (by 3.11) $H_h(\mathfrak{U})$ is of type $(\bar{\rho}, B)$ and therefore $\Gamma_{h-1}(\mathfrak{U}) \simeq 0$ mod B in A.

3.16. Let $\Gamma_h(\mathfrak{U})$ be a cycle mod B in A, where $B = B^1 \subset A = A^1 \subset M$. If $\Gamma_h(\mathfrak{U}) \sim 0 \bmod B$ in A, then $\rho\Gamma_h(\mathfrak{U}) \simeq 0 \bmod B$ in A. For if we have say $\phi X_{h+1}(\mathfrak{U}) = \Gamma_h(\mathfrak{U}) \bmod B$, $X_{h+1} \subset A$, then also $\phi X^i_{h+1}(\mathfrak{U}) = \Gamma^i_h(\mathfrak{U}) \bmod B$, hence $\phi\rho X_{h+1}(\mathfrak{U}) = \rho\Gamma_h(\mathfrak{U}) \bmod B$. $\rho X(\mathfrak{U})$ and $\rho\Gamma(\mathfrak{U})$ are both of type ρ, hence of type (ρ, B) and finally, $\rho X(\mathfrak{U}) \subset A$ since $A = A^1$.

3.17. Let $X(\mathfrak{B})$ be an arbitrary $\mathfrak{B}$-chain. If $\mathfrak{B}$ is a refinement of $\mathfrak{U}$ and if $\pi = \pi(\mathfrak{B}, \mathfrak{U})$, then on account of the invariance of π (2.20) we have $\pi X^i(\mathfrak{B}) = \pi(X(\mathfrak{B}))^i$, hence $\pi\rho X(\mathfrak{B}) = \rho\pi X(\mathfrak{B})$. Thus a $\mathfrak{B}$-chain of type ρ is projected by π into a $\mathfrak{U}$-chain of type ρ. Similarly, a $\mathfrak{B}$-cycle $\Gamma_h(\mathfrak{B}) \bmod B$ in A of type (ρ, B) is carried by π into a $\mathfrak{U}$-cycle of the same type; and if $\Gamma_h(\mathfrak{B}) \simeq 0 \bmod B$ in A, then the same holds for $\pi\Gamma_h(\mathfrak{B})$.

3.18. Let Γ_h be a cycle mod B in A where $0 \subseteq B = B^1 \subset A$. We shall say that Γ_h is of type (ρ, B) provided that

(i) each $\Gamma_h(\mathfrak{U})$ is of type (ρ, B); (ii) if $\mathfrak{B} \subset \mathfrak{U}$ and $\pi = \pi(\mathfrak{B}, \mathfrak{U})$, then $\pi\Gamma_h(\mathfrak{B}) \simeq \Gamma_h(\mathfrak{U}) \bmod B$ in A. These conditions being satisfied, we shall write $\Gamma_h \simeq 0 \bmod B$ in A provided that this relation holds for each $\Gamma_h(\mathfrak{U})$.

3.19. We shall consistently denote by $\mathfrak{p}$ the field of integers reduced modulo p. *Let* Γ_h *be a cycle* mod L *of type* (ρ, L), *say* $\Gamma_h(\mathfrak{U}) = \rho X_h(\mathfrak{U}) \bmod L$. *Let* $\Lambda X_h(\mathfrak{U}) = X^*(\mathfrak{U}^*)$. *If* $\rho = \sigma$ *or if* $\mathfrak{f} = \mathfrak{p}$ *and* $\rho = \bar{\sigma}$, $X^* = \{X^*(\mathfrak{U}^*)\}$ *is a cycle* mod L^* *in* M^*. *If* $\Gamma_h \simeq 0 \bmod L$, *then* $X^* \sim 0 \bmod L^*$.

Proof. $\phi X_h(\mathfrak{U})$ is a cycle of type $(\bar{\rho}, L)$ (3.14), say $\phi X_h(\mathfrak{U}) = \bar{\rho} Y_{h-1}(\mathfrak{U}) \bmod L$. Hence $\phi\Lambda X_h(\mathfrak{U}) = \Lambda\phi X_h(\mathfrak{U}) = \Lambda\bar{\rho}Y_{h-1}(\mathfrak{U}) \bmod L$. This last chain equals $0 \bmod L$ if $\bar{\rho} = \bar{\sigma}$ and equals $p\Lambda Y(\mathfrak{U}) = 0 \bmod p$ if $\bar{\rho} = \sigma$. Hence $\Lambda X_h(\mathfrak{U})$ is a $\mathfrak{U}^*$-cycle ($\mathfrak{U}^* = \Lambda\mathfrak{U}$). Suppose that $\mathfrak{B}^* \subset \mathfrak{U}^*$, $\mathfrak{B}^* = \Lambda\mathfrak{B}$. By 2.21, there exists a $\mathfrak{B}_0 \subset \mathfrak{U}$ such that $\Lambda\mathfrak{B}_0 = \mathfrak{B}^*$. Let $\pi = \pi(\mathfrak{B}_0, \mathfrak{U})$. Then $\pi\Gamma_h(\mathfrak{B}_0) \simeq \Gamma(\mathfrak{U}) \bmod L$; hence there exists a chain $\rho X_{h+1}(\mathfrak{U})$ such that

$$\phi\rho X_{h+1}(\mathfrak{U}) - \pi\Gamma_h(\mathfrak{B}_0) + \Gamma(\mathfrak{U}) = 0 \bmod L.$$

Let

$$Z_h(\mathfrak{U}) = \phi X_{h+1}(\mathfrak{U}) - \pi X_h(\mathfrak{B}_0) - X_h(\mathfrak{U}).$$

We have $\rho Z_h(\mathfrak{U}) = 0 \bmod L$, hence $Z_h(\mathfrak{U}) = \bar{\rho}Z'(\mathfrak{U}) \bmod L$ (3.11) so that $\Lambda Z_h(\mathfrak{U}) = 0 \bmod L^*$. Therefore

$$0 = \Lambda\phi X_{h+1}(\mathfrak{U}) - \pi_\lambda \Lambda X_h(\mathfrak{B}_0) - \Lambda X_h(\mathfrak{U}) \bmod L^*,$$

that is, $\pi_\lambda X^*(\mathfrak{B}^*) \sim X^*(\mathfrak{U}^*) \bmod L^*$, which proves that X^* is a cycle mod L^*. The second part of the theorem is proved by much the same sort of reasoning.

3.20. *Let* $X^* = \{X^*(\mathfrak{U}^*)\}$ *be a cycle* mod L^* *in* M^* *and let* $\Gamma_h = \{\Gamma_h(\mathfrak{U})\} = \{\Lambda^{-1}X^*(\mathfrak{U}^*)\}$. Γ_h *is a cycle* mod L *of type* σ. *If* $X^* \sim 0 \bmod L^*$, *then* $\Gamma_h \sim 0 \bmod L$.

The proof offers no difficulties and will be omitted (see 3.8).

3.21. *Let* Γ_h *be a cycle* mod B *of type* (ρ, B). *For every* $\mathfrak{U}$ *let there be chosen a cycle* $\Gamma_{h-1}(\mathfrak{U})$ mod B *of type* $(\bar{\rho}, B)$ *such that* $\Gamma_h(\mathfrak{U}):\Gamma_{h-1}(\mathfrak{U})$ (3.14). *Then* $\Gamma_{h-1} = \{\Gamma_{h-1}(\mathfrak{U})\}$ *is a cycle* mod B *of type* $(\bar{\rho}, B)$.

PROOF. We may write $\Gamma_h(\mathfrak{U}) = \rho X_h(\mathfrak{U})$ mod B, $\phi X_h(\mathfrak{U}) = \Gamma_{h-1}(\mathfrak{U})$. We have to show that if $\pi = \pi(\mathfrak{V}, \mathfrak{U})$, $\mathfrak{V} \subset \mathfrak{U}$, then $\pi\Gamma_{h-1}(\mathfrak{V}) \simeq \Gamma_{h-1}(\mathfrak{U})$ mod B. We have $\pi\rho X_h(\mathfrak{V}) - \rho X_h(\mathfrak{U}) \simeq 0$ mod B. Hence since π is invariant, $\rho(\pi X(\mathfrak{V}) - X(\mathfrak{U}) \simeq 0$ mod B, and therefore by 3.15, $\phi(\pi X(\mathfrak{V}) - X(\mathfrak{U})) \simeq 0$ mod B. Hence $\pi\phi X(\mathfrak{V}) \simeq \phi X(\mathfrak{U})$ mod B, i.e. $\pi\Gamma_{h-1}(\mathfrak{V}) — \Gamma_{h-1}(\mathfrak{U})$ mod B.

It will be natural to indicate the relation which exists between the cycles Γ_h and Γ_{h-1} by writing $\Gamma_h:\Gamma_{h-1}$. Obviously this relation is additive—that is, if also $\Gamma_h':\Gamma_{h-1}'$, then $(\Gamma_h + \Gamma_h'):(\Gamma_{h-1} + \Gamma_{h-1}')$.

3.22. We shall now make certain assumptions concerning the homology characters of M. It will be convenient first to adopt the convention that a $(0, \mathfrak{U})$-chain is an absolute $(0, \mathfrak{U})$-cycle if and only if the sum of its coefficients if zero. Then, for example, the boundary of every $(1, \mathfrak{U})$-chain is an absolute $(0, \mathfrak{U})$-cycle, whereas a $(0, \mathfrak{U})$-chain consisting of a single vertex with a non-zero coefficient is not. Moreover, the 0-dimensional Betti number of M no longer equals the number β of quasi-components of M but is one less than β; in particular if M is connected, $B_0(M) = 0$ and conversely (see [1], p. 169).

3.23. We now assume that

III. For every neighborhood $A(a)$ $(a \subset M)$ there is a neighborhood $A_1(a) \subset A(a)$ such that every h-cycle $(0 \leqq h \leqq n)$ in $M - A$ is ~ 0 in $M - A_1$.

3.24. *Let* $A(a)$ *be a neighborhood of* a, $a \subset L$. *There exists an invariant* $A_1(a) \subset A(a)$ *such that for every (absolute) cycle* Γ_h *of type* ρ *in* $M - A$ *and with* $0 \leqq h \leqq n - 1$, *there exists a cycle* Γ_h' *of type* ρ *in* $M - A$ *and a cycle* Γ_{h+1} *of type* $\bar{\rho}$ *in* $M - A_1$ *such that* $\Gamma_{h+1}:\Gamma_h'$ *and* $\Gamma_h' \simeq \Gamma_h$ *in* $M - A$.

The proof will occupy the next four sections.

3.25. Since $h + 1 \leqq n$, there exists by III (3.23) a neighborhood B of a (with $B \subset A$) such that every h-cycle in $M - B$ is ~ 0 in $M - A$. Let A_0 be an invariant neighborhood of a such that $A_0 \subset B$. Obviously, h-cycles in $M - A$ are ~ 0 in $M - A_0$. Similarly choose an invariant neighborhood $A_1 \subset A_0$ such that $(h + 1)$-cycles in $M - A_0$ are ~ 0 in $M - A_1$.

Let $\mathfrak{U}$ be a definitely chosen covering and let $\mathfrak{U}_1 \subset \mathfrak{U}$. Let π_1 and π_{11} be two projections of $\mathfrak{U}_1$ into $\mathfrak{U}$. There exists an $(h + 1)$-chain $X(\mathfrak{U}_1)$ in $M - A_0$ such that $\phi X(\mathfrak{U}_1) = \Gamma(\mathfrak{U}_1)$. We assert that there exists in $M - A_0$ an h-chain $D(\mathfrak{U})$ of type ρ such that

$$\pi_1 X(\mathfrak{U}_1) - \pi_{11} X(\mathfrak{U}_1) - D(\mathfrak{U}) \sim 0 \text{ in } M - A_0 .$$

In fact, relative to π_1 and π_{11} there is determined ([1], p. 159) for each $(k, \mathfrak{U}_1)$-chain $Y(\mathfrak{U}_1)$ in a given set H say, a $(k + 1, \mathfrak{U})$-chain $P[Y(\mathfrak{U}_1)]$ in H which, as a

function of $Y(\mathfrak{U}_1)$, is additive and satisfies the relation (written here for the $(h + 1)$-chain $X(\mathfrak{U}_1)$):

$$(1) \qquad \phi P[X(\mathfrak{U}_1)] = \pi_1 X(\mathfrak{U}_1) - \pi_{11} X(\mathfrak{U}_1) - P[\phi X(\mathfrak{U}_1)].$$

Let $D(\mathfrak{U}) = P[\phi X(\mathfrak{U}_1)] = P[\Gamma_h(\mathfrak{U}_1)]$. Since $\Gamma_h(\mathfrak{U}_1)$ is of type ρ and P is additive, it follows that $D(\mathfrak{U})$ is of type ρ. Moreover $D(\mathfrak{U}_1) \subset M - A_0$ since $\Gamma(\mathfrak{U}_1) \subset M - A \subset M - A_0$. Therefore (1) determines a homology of the desired form.

3.26. Now let $\mathfrak{U}_1$ and $\mathfrak{U}_2$ be normal refinements of $\mathfrak{U}$ relative to cycles in $M - A_1$. Let $\pi_i = \pi_i(\mathfrak{U}_i, \mathfrak{U})$ $(i = 1, 2)$. There exists in $M - A_0$ an $(h + 1)$-chain $X(\mathfrak{U}_i)$ bounded by $\Gamma_h(\mathfrak{U}_i)$. We assert that there exists in $M - A_1$ a chain $E(\mathfrak{U})$ of type ρ such that

$$(1) \qquad \pi_1 X(\mathfrak{U}_1) - \pi_2 X(\mathfrak{U}_2) - E(\mathfrak{U}) \sim 0 \text{ in } M - A_1 .$$

For, let $\mathfrak{W} \subset \mathfrak{U}_1 \mathfrak{U}_2$ and let $\pi_{wi} = \pi_{wi}(\mathfrak{W}, \mathfrak{U}_i)$ $(i = 1, 2)$. Since Γ_h is of type ρ and is in $M - A_0$, there exist (by 3.18) chains $G(\mathfrak{U}_i)$ of type ρ in $M - A_0$ such that

$$(2) \qquad \phi G(\mathfrak{U}_i) = \pi_{wi} \Gamma_h(\mathfrak{W}) - \Gamma_h(\mathfrak{U}_i) \qquad (i = 1, 2).$$

Moreover, by the choice of A_0 there exists in $M - A_0$ a chain $X(\mathfrak{W})$ bounded by $\Gamma_h(\mathfrak{W})$. From (2) we see that

$$\pi_{wi} X(\mathfrak{W}) - X(\mathfrak{U}_i) - G(\mathfrak{U}_i)$$

is a cycle in $M - A_0$; its projection by π_i is

$$Q_i(\mathfrak{U}) = \pi_i \pi_{wi} X(\mathfrak{W}) - \pi_i X(\mathfrak{U}_i) - \pi_i \phi G(\mathfrak{U}_i) \qquad (i = 1, 2)$$

and this cycle is still in $M - A_0$ and it is essential. Hence there exists in $M - A_0$ an $(h + 1)$-cycle of which $Q_i(\mathfrak{U})$ is a "coefficient" (cf. footnote 4a); that cycle, and hence $Q_i(\mathfrak{U})$, is ~ 0 in $M - A_1$. The same is true of $Q_1(\mathfrak{U}) - Q_2(\mathfrak{U})$—that is,

$$(3) \quad \pi_1 X(\mathfrak{U}_1) - \pi_2 X(\mathfrak{U}_2) - \pi_1 \pi_{w1} X(\mathfrak{W}) + \pi_2 \pi_{w2} X(\mathfrak{W}) \\ - Y(\mathfrak{U}) \sim 0 \text{ in } M - A_1,$$

where $Y(\mathfrak{U})$ is a chain of type ρ in $M - A_1$, hence in $M - A_0$. By 3.25, we may write

$$(4) \qquad -\pi_1 \pi_{w1} X(\mathfrak{W}) + \pi_2 \pi_{w2} X(\mathfrak{W}) - F(\mathfrak{U}) \sim 0 \text{ in } M - A_0$$

where $F(\mathfrak{U})$ is a chain if type ρ in $M - A_0$, hence in $M - A_1$. Since (4) also holds in $M - A_1$, a relation of the form (1) is obtained by subtracting (3) from (4).

3.27. Now let $\mathfrak{U}$ be arbitrary. For each $\mathfrak{U}$, choose a normal refinement $\mathfrak{U}_1$ relative to cycles in $M - A_1$, and a chain $X(\mathfrak{U}_1)$ in $M - A_0$ bounded by $\Gamma_h(\mathfrak{U}_1)$. For each $\mathfrak{U}_1$ and $\mathfrak{U}$ choose a $\pi_{1u} = \pi_{1u}(\mathfrak{U}_1, \mathfrak{U})$ and let

$$X'(\mathfrak{U}) = \pi_{1u}X(\mathfrak{U}_1), \qquad \Gamma'_h(\mathfrak{U}) = \pi_{1u}\Gamma_h(\mathfrak{U}_1).$$

We shall show that if $\mathfrak{B} \subset \mathfrak{U}$ and $\pi = \pi(\mathfrak{B}, \mathfrak{U})$, there exists in $M - A_1$ a chain $H(\mathfrak{U})$ of type ρ such that

$$\pi X'(\mathfrak{B}) - X'(\mathfrak{U}) - H(\mathfrak{U}) \sim 0 \text{ in } M - A_1. \tag{1}$$

By 3.26, there exists in $M - A_1$ a chain $H(\mathfrak{U})$ of type ρ such that

$$\pi\pi_{1v}X(\mathfrak{B}_1) - \pi_{1u}X(\mathfrak{U}_1) - H(\mathfrak{U}) \sim 0 \text{ in } M - A_1.$$

But this is precisely the relation we wish, since $\pi_{1v}X(\mathfrak{B}_1) = X'(\mathfrak{B})$ and $\pi_{1u}X(\mathfrak{U}_1) = X'(\mathfrak{U})$.

From 3.13, it follows that $\bar{\rho}X'(\mathfrak{U})$ is a $\mathfrak{U}$-cycle in $M - A_1$ of type $\bar{\rho}$. Let $\Gamma_{h+1} = \{\bar{\rho}X'(\mathfrak{U})\}$. We assert that Γ_{h+1} is a cycle of type $\bar{\rho}$ in $M - A_1$. The homology (1) holds when the chains involved are replaced by their images under T. Hence (3.16)

$$\pi\bar{\rho}X'(\mathfrak{B}) - \bar{\rho}X'(\mathfrak{U}) - \bar{\rho}H(\mathfrak{U}) \simeq 0 \text{ in } M - A_1. \tag{2}$$

But $\bar{\rho}H(\mathfrak{U}) = 0$ since $H(\mathfrak{U})$ is of type ρ. Consequently (2) may be written

$$\pi\Gamma_{h+1}(\mathfrak{B}) - \Gamma_{h+1}(\mathfrak{U}) \simeq 0 \text{ in } M - A_1$$

and our assertion is proved.

3.28. To complete the proof of 3.24 it is merely necessary to recall that $\phi X'(\mathfrak{U}) = \pi_{1u}\Gamma_h(\mathfrak{U}_1) = \Gamma'(\mathfrak{U}_1)$ so that $\Gamma_{h+1}:\Gamma'_h$. Moreover, $\pi_{1u}\Gamma_h(\mathfrak{U}_1) \simeq \Gamma_h(\mathfrak{U})$ in $M - A$ (3.18). Hence $\Gamma'_h \simeq \Gamma_h$ in $M - A$ and the proof is now complete.

3.29. Let Γ be a cycle mod B in A $(0 \subseteqq B = B' \subset A)$ of type (ρ, B). On account of the properties of special coverings and of invariant projections it is readily seen that $\{\Gamma^i(\mathfrak{U})\}$ is also a cycle mod B in A of type (ρ, B). It will be natural to denote this cycle by Γ^i.

3.30. Let $\mathfrak{r}$ be a field such that the relation $px = 0$ $(x \in \mathfrak{r})$ implies $x = 0$ and the relation $x^p = 1$ implies $x = 1$ if $p > 2$ and $x = \pm 1$ if $p = 2$. The field of rational numbers for example satisfies these conditions. We shall now assume that

IV. The coefficient field $\mathfrak{f}$ is either $\mathfrak{p}$ (see 3.19) or $\mathfrak{r}$.

V. The Betti numbers $B_k(M; \mathfrak{f})$ $(0 \leqq k \leqq n)$ of M relative to the coefficient field $\mathfrak{f}(= \mathfrak{p}$ or $\mathfrak{r})$ are the same as those of an n-sphere.

3.31. In particular we have $B_0(M; \mathfrak{f}) = 0$ (see 3.22) and $B_n(M; \mathfrak{f}) = 1$. From the second relation there exists a cycle[7] Δ_n not ~ 0 and such that every cycle Γ_n in M satisfies a homology of the form $\Gamma_n \sim x\Delta_n$, $x \in \mathfrak{f}$. In particular, $\Delta_n^1 \sim y\Delta_n (y \in \mathfrak{f})$. Hence $\Delta_n^2 \sim y\Delta_n^1 \sim y^2\Delta_n$ etc. so that $\Delta_n^p = \Delta_n \sim y^p\Delta_n$. Hence $y^p = 1$ and therefore (3.30) $y = 1$ unless $p = 2$ in which case $y = \pm 1$. Thus $\Delta_n^1 \sim \Delta_n$ when $p > 2$ and $\Delta_n^1 \sim \Delta_n$ or $-\Delta_n$ when $p = 2$.

Let ρ_n stand for σ when $p > 2$ and for σ or $\bar{\sigma}$ according as $\Delta_n^1 \sim \Delta_n$ or $-\Delta_n$ when $p = 2$. Regardless of the value of p we have $\bar{\rho}_n\Delta_n \sim 0$. Hence there exists for each $\mathfrak{U}$ an $(n + 1)$-chain $X_{n+1}(\mathfrak{U})$ bounded by $\bar{\rho}_n\Delta_n(\mathfrak{U})$. But since every $\mathfrak{U}$-simplex not in L is of dimension $\leqq n$ (2.16), we have $X_{n+1}(\mathfrak{U}) \subset L$ and hence $\phi X_{n+1}(\mathfrak{U}) \subset L$, i.e. $\bar{\rho}_n\Delta_n(\mathfrak{U}) = 0 \bmod L$ and therefore (3.11), $\Delta_n(\mathfrak{U})$ is of type (ρ, L), and, in fact, Δ_n itself is of type (ρ_n, L) (see 3.18).

Remarks: 1. From the preceding paragraph it is clear that a homology of the form $\Gamma_n(\mathfrak{U}) \sim 0 \bmod B$ implies $\Gamma_n(\mathfrak{U}) = 0 \bmod (B + L)$.

2. We have seen that regardless of the value of p we have $\bar{\rho}_n\Delta_n = 0 \bmod L$; it is easy to see also that $\rho_n\Delta_n = p\Delta_n \bmod L$.

3. If $\mathfrak{f} = \mathfrak{p}$, then regardless of the value of p, $\sigma\Delta_n = \bar{\sigma}\Delta_n = 0 \bmod L$, so that Δ_n is simultaneously of types (σ, L) and $(\bar{\sigma}, L)$. We shall agree however to take $\rho_n = \sigma$ when $\mathfrak{f} = \mathfrak{p}$.

3.32. Let us choose once and for all a family of chains $\{\chi_n(\mathfrak{U})\}$ such that $\Delta_n(\mathfrak{U}) = \rho_n\chi(\mathfrak{U}) \bmod L$. Let $\Delta_{n-1} = \{\phi\chi_n(\mathfrak{U})\}$. As we have seen (3.21) Δ_{n-1} is a cycle of type $(\bar{\rho}_n, L)$ and therefore we may write $\Delta_{n-1}(\mathfrak{U}) = \rho_{n-1}\chi_{n-1}(\mathfrak{U}) \bmod L$, where $\rho_{n-1} = \bar{\rho}_n$. Let $\Delta_{n-2} = \{\phi\chi_{n-1}(\mathfrak{U})\}$ and continue in this manner down to the dimension 0. We obtain thus a sequence of absolute cycles $\Delta_n, \cdots, \Delta_0$ such that

$$\Delta_h = \rho_h\chi_h \bmod L \qquad (h = n, \cdots, 0)$$

$$\phi\chi_h = \Delta_{h-1} \qquad (h = n, \cdots, 1)$$

$$\Delta_h : \Delta_{h-1} \qquad (h = n, \cdots, 1)$$

where $\rho_h = \rho_n$ or $\bar{\rho}_n$ according as $n - h$ is even or odd.

3.33. *Let $X(\mathfrak{U})$ be a chain of type ρ. If $\rho = \bar{\sigma}$ or if $\rho = \sigma$ and $\mathfrak{f} = \mathfrak{p}$, then $L \cap X(\mathfrak{U}) = 0$ (i.e. no simplex of $X(\mathfrak{U})$ is in L).*

Proof. Let $X(\mathfrak{U}) = \rho Y(\mathfrak{U})$, and let $Y(\mathfrak{U}) = Y_0 + Y_{00}$ where $Y_0 = L \cap X(\mathfrak{U})$. Obviously $Y_0 \subset L$, $Y_{00} \cap L = 0$, $Y_{00}^i \cap L = 0$. Hence our assertion follows from the relations

$$X(\mathfrak{U}) = \bar{\sigma}Y(\mathfrak{U}) = \bar{\sigma}Y_{00} \quad \text{when} \quad \rho = \bar{\sigma};$$

$$X(\mathfrak{U}) = \sigma Y(\mathfrak{U}) = pY_0 + \sigma Y_{00} = \sigma Y_{00} \quad \text{when} \quad \rho = \sigma, \mathfrak{f} = \mathfrak{p}.$$

---

[7] Δ_n of course depends on $\mathfrak{f}$; it will be understood however that statements about Δ_n are true for either choice of $\mathfrak{f}$ unless it is stated to the contrary.

3.34. *Let $X(\mathfrak{U})$ be a chain of type $(\rho, B + L)$ where $0 \subsetneq B = B^1$. If $\rho = \bar{\sigma}$ or if $\mathfrak{f} = \mathfrak{p}$ and $\rho = \sigma$, $X(\mathfrak{U})$ is of type (ρ, B).*

PROOF. By hypotheses $X(\mathfrak{U}) = \rho Y(\mathfrak{U}) \bmod (B + L)$; hence by 3.33, $X(\mathfrak{U}) = \rho Y(\mathfrak{U}) \bmod B$.

3.35. *Let Γ be a cycle* mod $(B + L)$ *of type $(\rho, B + L)$ where $0 \subsetneq B = B^1$. If $\rho = \bar{\sigma}$ or if $\mathfrak{f} = \mathfrak{p}$ and $\rho = \sigma$, Γ is a cycle* mod B *of type (ρ, B). If $\Gamma \simeq 0$* mod *then also $\Gamma \simeq 0$* mod B. *These statements hold for $\mathfrak{U}$-cycles as well as cycles.*

PROOF. From 3.34, $\Gamma(\mathfrak{U})$ is a chain of type (ρ, B). Hence we have say $\Gamma(\mathfrak{U}) = \rho X(\mathfrak{U}) \bmod B$. Now $\phi\Gamma(\mathfrak{U}) = 0 \bmod (B + L)$. But $\phi\Gamma(\mathfrak{U}) = \phi\rho X(\mathfrak{U}) \bmod (B + L)$ and hence (3.33) $\phi\Gamma(\mathfrak{U}) = 0 \bmod B$ so that $\Gamma(\mathfrak{U})$ is a cycle mod B of type (ρ, B). Another application of 3.33 shows readily that if $\mathfrak{B} \subset \mathfrak{U}$ and $\pi = \pi(\mathfrak{B}, \mathfrak{U})$, the relation $\pi\Gamma(\mathfrak{B}) \simeq \Gamma(\mathfrak{U}) \bmod (B + L)$ holds also mod B. The proof of the second part of the theorems is similar.

3.36. *L is nowhere dense in M.*

PROOF. Let $\mathfrak{f} = \mathfrak{p}$. Then Δ_n is of type (σ, L) (3.31) hence by 3.35, Δ_n is of type σ and $\Delta_n \cap L = 0$. Suppose L contains an open set A. Then since $\Delta_n \subset M - L \subset M - A$, we have by assumption III, $\Delta_n \sim 0$ which is impossible.

This result is similar to a theorem of Newman (a corollary of the more general theorem of [10]) which asserts that if a periodic transformation T operates in a locally euclidean space and leaves fixed the points of an open set, T must be the identity. Our result holds, incidentally, regardless of the values of the Betti numbers of M of dimension $< n$ since only the existence of Δ_n was used in the proof.

3.37. *Δ_n can not be ~ 0* mod L.

PROOF. Suppose $\Delta_n \sim 0 \bmod L$. Then (see 3.31) $\Delta_n(\mathfrak{U}) = 0 \bmod L$ for every $\mathfrak{U}$, hence $\Delta_n \subset L = M - (M - L)$. Since T is not the identity, the open set $M - L$ is non-vacuous and it follows from III (3.23) that $\Delta_n \sim 0$ which is impossible.

3.38. *The open set $M - L$ has at most two components. If $p > 2$, $M - L$ is connected. If $p = 2$ and if $M - L$ has two components, they are images of each other under T; moreover we have in this case $\Delta_n^1 \sim -\Delta_n$.*

PROOF. Suppose $M - L$ has two or more components and that one of them is A. Then $A, A^1, \cdots, A^{p-1}$ are components of $M - L$ and it is clear that either these sets are mutually exclusive or else $A = A^1 = \cdots = A^{p-1}$. If the latter possibility holds, there must be at least one other component say B. say B. Let T_0 be a transformation of M into itself defined as follows: T_0 is to leave the points of A fixed but elsewhere is to be identical with T. Obviously T_0 is not the identity; it is however of period p and its invariant set contains an open set. Since this contradicts 3.36, we conclude that $A, A^1, \cdots, A^{p-1}$ are mutually exclusive. Suppose there are still other components of $M - L$. We

could then again define a transformation of period p, different from the identity and this time leaving fixed each point of σA, which is impossible. Hence the components of $M - L$ are precisely $A, A_1, \cdots, A_p$. We shall show that $p = 2$.

Suppose on the contrary that $p \geqq 3$. Let $H \neq 0$ be an open set such that $\bar{H} \subset A$. Since A, A^1, A^2 are mutually exclusive and $\bar{H}^i \subset A^i$, it is easy to see that every $\mathfrak{U}$ has a special refinement $\mathfrak{U}_3$ such that (i) $\Delta_n(\mathfrak{U}_3)$ is not $\sim 0 \bmod L$ (see 3.37); (ii) no $\mathfrak{U}_3$-simplex in H can be in A^1 and (iii) no $\mathfrak{U}_3$-simplex in $A + A^1$ can be in H^2. Let $\mathfrak{U}_2$ be a normal refinement of $\mathfrak{U}_3$ relative to absolute cycles in $M - H$ and absolute cycles in $M - H^2$. There exists a refinement $\mathfrak{U}_1$ of $\mathfrak{U}_2$ ($\mathfrak{U}_1$ not necessarily special) such that every $\mathfrak{U}_2$-chain simultaneously in $\bar{A}$ and in $\overline{M - A}$ is also in $\bar{A}(\overline{M - A})$.[8] Let $\mathfrak{U}_0$ be a *special* refinement of $\mathfrak{U}_1$. Let $D(\mathfrak{U}_0) = A \cap \Delta_n(\mathfrak{U}_0)$, $C(\mathfrak{U}_0) = \Delta_n(\mathfrak{U}_0) - D(\mathfrak{U}_0)$. Then

$$D(\mathfrak{U}_0) \subset A \subset \bar{A} \tag{1}$$

$$C(\mathfrak{U}_0) \subset M - A \subset \overline{M - A} \tag{2}$$

$$C(\mathfrak{U}_0) + D(\mathfrak{U}_0) = \Delta_n(\mathfrak{U}_0).$$

Let $\pi_i = \pi_i(\mathfrak{U}_{i-1}, \mathfrak{U}_i)$[9] and let

$$^{(i)}\Delta_n(\mathfrak{U}_i) = \pi_i\Delta_n(\mathfrak{U}_0), \qquad ^{(i)}D(\mathfrak{U}_i) = \pi_i D_n(\mathfrak{U}_0), \qquad ^{(i)}C(\mathfrak{U}_i) = \pi_i C(\mathfrak{U}_0).$$

The chains $^{(1)}D(\mathfrak{U}_1)$, $^{(1)}C(\mathfrak{U}_1)$ satisfy (1) and (2) respectively, hence so do their boundaries. Since $\phi^{(1)}D(\mathfrak{U}_1) = -\phi^{(1)}C(\mathfrak{U}_1)$, we have, on account of the choice of $\mathfrak{U}_1$,

$$\phi^{(1)}D(\mathfrak{U}_1) \subset \bar{A}(\overline{M - A}) \subset L.$$

This relation holds also for $^{(2)}D(\mathfrak{U}_2)$ and since $\mathfrak{U}_2$ is special, it follows that $\phi^{(2)}D(\mathfrak{U}_2)$ is invariant (3.5), hence $\bar{\sigma}^{(2)}D(\mathfrak{U}_2)$ is an absolute $(n, \mathfrak{U}_2)$-cycle. Since π_3 is invariant, we have $\pi_3\bar{\sigma}^{(3)}D(\mathfrak{U}_2) = \bar{\sigma}^{(3)}D(\mathfrak{U}_3)$ and therefore the cycle $\bar{\sigma}^{(3)}D(\mathfrak{U}_3)$ is essential. Since $\bar{\sigma}^{(3)}D(\mathfrak{U}_3) \subset A + A^1$ and since therefore $\bar{\sigma}^{(3)}D(\mathfrak{U}_3) \subset M - H^2$ (by iii), it follows that $\bar{\sigma}^{(3)}D(\mathfrak{U}_3) \sim 0$. This implies that $\bar{\sigma}^{(3)}D(\mathfrak{U}_3) = 0 \bmod L$ (3.31 remark 1), therefore $\bar{\sigma}^{(3)}D(\mathfrak{U}_3) = 0$ (3.33). Thus $^{(3)}D(\mathfrak{U}_3) = {}^{(3)}D^1(\mathfrak{U}_3)$. Now $^{(3)}D(\mathfrak{U}_3) \cap H \neq 0$; for otherwise, since $^{(3)}C(\mathfrak{U}_3) \subset M - A \subset M - H$, we would have $^{(3)}\Delta_n(\mathfrak{U}_3) = {}^{(3)}C(\mathfrak{U}_3) + {}^{(3)}D(\mathfrak{U}_3) \subset M - H$, and $^{(3)}\Delta_n(\mathfrak{U}_3)$ being essential would therefore be ~ 0; hence $\Delta_n(\mathfrak{U}_3) \sim 0$ which would contradict (i). Thus some simplex of $^{(3)}D(\mathfrak{U}_3)$ is in H. But by (ii) this is impossible, since $D(\mathfrak{U}_3) = {}^{(3)}D^1(\mathfrak{U}_3) \subset A^1$. Thus p must equal 2. We shall show that now $\Delta_n^1 \sim -\Delta_n$. As we have seen, the cycle $\bar{\sigma}^{(3)}D(\mathfrak{U}_3)$ is essential and therefore is $\sim x\Delta_n(\mathfrak{U}_3) \sim x^{(3)}\Delta_n(\mathfrak{U}_3)$, and we can not have $x = 0$, otherwise we arrive at a contradiction as above. Now we have $(\bar{\sigma}^{(3)}D(\mathfrak{U}_3))^1 \sim x\Delta^1(\mathfrak{U}_3)$. But since

[8] On account of I (2.13), M is perfectly normal ([2], p. 40). Hence M is completely normal ([2], p. 40) and therefore ([5], p. 8) a refinement such as $\mathfrak{U}_1$ exists.

[9] π_1 and π_2 are not necessarily invariant.

$p = 2$, $(\bar{\sigma}^3 D(\mathfrak{U}_3))^1 = -\bar{\sigma}^{(3)} D(\mathfrak{U}_3) \sim -x\Delta_n(\mathfrak{U}_3)$. Hence, taking $\mathfrak{f} = \mathfrak{r}$, it follows that $\Delta_n^1(\mathfrak{U}_3) \sim -\Delta_n(\mathfrak{U}_3)$ and since $\mathfrak{U}_3$ is a refinement of $\mathfrak{U}$ and $\mathfrak{U}$ is arbitrary, it follows that $\Delta_n^1 \sim -\Delta_n$. This relation holds automatically when $\mathfrak{f} = \mathfrak{p}$.

3.39. *Let Γ_h be a cycle of type ρ with $0 \leqq h \leqq n - 1$; in particular if $h = n - 1$, assume that Γ_h is not $\simeq 0 \bmod L$.*[10] *There exist cycles Γ_{h+1} and Γ_h' of types $\bar{\rho}$ and ρ such that $\Gamma_{h+1}:\Gamma_h'$ and $\Gamma_h' \simeq \Gamma_h$.*

PROOF. If $h \leqq n - 1$, the proof is the same as that of 3.24 except that now we take $A = A^1 = M$ and make use of V (3.30) instead of III (3.23).

Assume then that $h = n - 1$. On account of the hypothesis on Γ_{n-1}, there exists a covering say $\mathfrak{W}$ such that $\Gamma_{n-1}(\mathfrak{W})$ is not $\simeq 0 \bmod L$. Now choose for every $\mathfrak{U}$ a normal refinement $\mathfrak{U}_0$ relative to absolute cycles and let $\mathfrak{U}_1 \subset \mathfrak{U}_0\mathfrak{W}$. Clearly $\Gamma_{n-1}(\mathfrak{U}_0)$ is not $\simeq 0 \bmod L$.

Now $\mathfrak{U}_1$ like $\mathfrak{U}_0$ is a normal refinement of $\mathfrak{U}$ and therefore, if for each $\mathfrak{U}_1$ we choose an n-chain $X(\mathfrak{U}_1)$ bounded by $\Gamma_{n-1}(\mathfrak{U}_1)$ (which we may do by V (3.30)) and then let $G_n = \bar{\rho}X_n'(\mathfrak{U}_1) =$ a $\mathfrak{U}_1$-cycle of type $\bar{\rho}$, the cycle $\Gamma_n(\mathfrak{U}) = \pi_1 G(\mathfrak{U}_1)$ ($\pi_1 = \pi_1(\mathfrak{U}_1, \mathfrak{U})$) is essential. Consequently we may write

$$\bar{\rho}X_n(\mathfrak{U}) = \Gamma_n(\mathfrak{U}) \sim x_u\Delta_n(\mathfrak{U}) \qquad (x_u \in \mathfrak{f},\ X_n(\mathfrak{U}) = \pi_1 X'(\mathfrak{U}))$$

from which it follows (3.31) that

$$\Gamma_n(\mathfrak{U}) = x_u\Delta_n(\mathfrak{U}) \bmod L. \tag{1}$$

Now suppose that $\bar{\rho} = \rho_n$. Then (1) takes the form

$$\rho_n X_n(\mathfrak{U}) = x_u\rho_n\chi_n(\mathfrak{U}) \simeq x_u\rho_n\chi_n(\mathfrak{U}) \bmod L. \tag{2}$$

Hence by 3.15,

$$\Gamma_{n-1}(\mathfrak{U}) \simeq x_u\Delta_{n-1}(\mathfrak{U}) \bmod L. \tag{3}$$

In the same way, if $\mathfrak{V}$ is a second covering we have

$$\Gamma_{n-1}(\mathfrak{V}) \simeq x_v\Delta_{n-1}(\mathfrak{V}) \bmod L. \tag{4}$$

Suppose $\mathfrak{V} \subset \mathfrak{U}$, $\pi = \pi(\mathfrak{V}, \mathfrak{U})$. Then on taking the projection of (4) we obtain (see 3.17)

$$\pi\Gamma_{n-1}(\mathfrak{V}) \simeq \pi x_v\Delta_{n-1}(\mathfrak{V}) \simeq x_v\Delta_{n-1}(\mathfrak{U}) \bmod L.$$

Since $\pi\Gamma_{n-1}(\mathfrak{V}) \simeq \Gamma_{n-1}(\mathfrak{U})$, we obtain from (3),

$$(x_v - x_u)\Gamma_{n-1}(\mathfrak{U}) \simeq 0 \bmod L,$$

so that $x_v = x_u$, since $\Gamma_{n-1}(\mathfrak{U})$ is not $\simeq 0$. In case $\mathfrak{V}$ is not a refinement of $\mathfrak{U}$ we still have $x_v = x_u$ since both numbers will equal x_w (say) corresponding to a common refinement $\mathfrak{W}$ of $\mathfrak{U}$, $\mathfrak{V}$. It follows that the relation (1) may now be written $\Gamma_n(\mathfrak{U}) = x\Delta(\mathfrak{U})$ where x is independent of $\mathfrak{U}$ and therefore $\Gamma_n = \{\Gamma_n(\mathfrak{U})\}$

---

[10] Since Γ_h is of type ρ, it is obviously of type (ρ, L).

is a cycle since it equals $x\Delta_n$. If we let $\phi X_n(\mathfrak{U}) = \Gamma'_{n-1}(\mathfrak{U})$, we evidently have $\Gamma_n : \Gamma'_{n-1} \simeq \Gamma_{n-1} \bmod L$.

We turn now to the case $\rho = \rho_n$. This can only occur if $\mathfrak{f} = \mathfrak{p}$. For suppose $\mathfrak{f} = \mathfrak{r}$. In any case we have (2) with ρ_n replaced by $\bar{\rho}_n$ and therefore, since $\rho_n \bar{\rho}_n \chi_n(\mathfrak{U}) = 0$,

$$0 = \rho_n x_u \Delta_n(\mathfrak{U}) = p x_u \Delta_n(\mathfrak{U}) \bmod L$$

for every $\mathfrak{U}$. Hence by 3.37, $px_u = 0$, hence $x_u = 0$. Consequently from (1) we have $\Gamma_n \simeq 0 \bmod L$ which is impossible.

Suppose then that $\rho = \rho_n$, $\mathfrak{f} = \mathfrak{p}$. Since now Δ_n is of type $(\bar{\rho}_n, L)$ (as well as (ρ_n, L) (see 3.32, remark 3)), we may write $\Delta_n(\mathfrak{U}) = \bar{\rho}_n \chi'_n(\mathfrak{U}) \bmod L$, $\Delta'_{n-1}(\mathfrak{U}) = \phi\chi'_n(\mathfrak{U})$ and the argument then proceeds precisely as in the case $\bar{\rho} = \rho_n$, using now Δ', χ' instead of Δ, χ, and replacing ρ_n by $\bar{\rho}_n$.

3.40. *If* Γ_h $(0 \leqq h \leqq n)$ *is a cycle of type* ρ_h, *then* $\Gamma_h \simeq x\Delta_h \bmod L$, $x \in \mathfrak{f}$.

PROOF. Let $h = n$. Then we have $\Gamma_n \sim x\Delta_n (x \in \mathfrak{f})$. This implies $\Gamma_n = x\Delta_n \bmod L$, hence $\Gamma_n \simeq x\Delta_n \bmod L$.

Suppose next that the theorem is true for the dimension $h + 1 \leqq n$. By 3.39, there exist cycles Γ_{h+1}, Γ'_h of types $\bar{\rho}_h = \rho_{h+1}$ and ρ_h respectively such that $\Gamma_{h+1} : \Gamma'_h \simeq \Gamma_h$. By the hypothesis of the induction we have $\Gamma_{h+1} - y\Delta_{h+1} \simeq 0 \bmod L$ and since $(\Gamma_{h+1} - y\Delta_{h+1}) : (\Gamma'_h - y\Delta_h)$ it follows from 3.15 that $\Gamma_h - y\Delta_h \simeq \Gamma'_h - y\Delta_h \simeq 0 \bmod L$.

3.41. *If* $\mathfrak{f} = \mathfrak{r}$ *and if* Γ_h *is a cycle of type* $\bar{\rho}_h$ $(0 \leqq h \leqq n)$ *then* $\Gamma_h \simeq 0 \bmod L$.

PROOF. Suppose $h = n$. We have $\Gamma_h = x\Delta_n \bmod L$ and since $\rho_n \Gamma_n = 0$, $\rho_n \Delta_n = p\Delta_n$, it follows that $px\Delta_n = 0 \bmod L$, hence (3.37) $px = 0$, $x = 0$. Thus $\Gamma_n \simeq 0 \bmod L$. The proof now proceeds by induction precisely as the proof of 3.40.

3.42. *Let* $A \neq 0$ *be an open set in* M. *Then* Δ_n *is not* $\sim 0 \bmod (M - A)$.

PROOF. Let $\Delta_n(\mathfrak{U}) = \Delta'(\mathfrak{U}) + \Delta''(\mathfrak{U})$ where $\mathfrak{U}$ is arbitrary and $\Delta'' = (M - A) \cap \Delta_n(\mathfrak{U})$. Suppose contrary to hypothesis that there is an $X_{n+1}(\mathfrak{U})$ such that $\phi X_{n+1}(\mathfrak{U}) = \Delta(\mathfrak{U}) = \Delta'(\mathfrak{U}) \bmod (M - A)$. We may then write

$$\phi X(\mathfrak{U}) = \Delta'(\mathfrak{U}) + Y(\mathfrak{U}) \sim 0 \tag{1}$$

where $Y(\mathfrak{U}) \subset M - A$. Now $Y(\mathfrak{U}) - \Delta''(\mathfrak{U})$ is an absolute cycle since

$$\phi(Y - \Delta'') = \phi(Y - \Delta + \Delta') = \phi(Y + \Delta') = \phi\phi X = 0$$

and moreover $Y(\mathfrak{U}) - \Delta''(\mathfrak{U}) \subset M - A$. We may in fact assume by the usual argument (cf. for example the proof of 3.39) that $Y(\mathfrak{U}) - \Delta''(\mathfrak{U})$ is essential in $M - A$; hence by III (3.23), $Y(\mathfrak{U}) - \Delta''(\mathfrak{U}) \sim 0$ and on adding this homology to (1) we obtain $\Delta_n \sim 0$ which is impossible.

3.43. *Let $A(a)$ be a neighborhood of a, $a \subset M$, and let Γ_h $(0 \leqq h \leqq n - 1)$ be a cycle* mod $(M - A)$. *There exists an $A_1(a) \subset A(a)$ such that $\Gamma_h \sim 0$* mod $(M - A_1)$.

PROOF. From 1.2, $\phi\Gamma_h$ is a cycle in $M - A$. From III (3.23), there exists an $A_1(a) \subset A(a)$ such that $\phi\Gamma_h \sim 0$ in $M - A_1$, say $\phi Y_h(\mathfrak{U}) = \phi\Gamma_h(\mathfrak{U})$, $Y_h(\mathfrak{U}) \subset M - A_1$. By the usual argument we may assume that the cycle $Y_h(\mathfrak{U}) - \Gamma_h(\mathfrak{U})$ is essential; hence, since $h \leqq n - 1$, $Y_h(\mathfrak{U}) - \Gamma_h(\mathfrak{U}) \sim 0$. Hence $\Gamma_h(\mathfrak{U}) \sim 0$ mod $(M - A_1)$ for every $\mathfrak{U}$.

3.44. Let $\beta_k(a, M; \mathfrak{f})$ be the k-dimensional local Betti number of M at the point a, relative to the coefficient field $\mathfrak{f}$. It follows immediately from the definition of the local Betti numbers ([4], page 680) and from 3.43, that $\beta_k(a, M; \mathfrak{f}) = 0$ when $0 \leqq k \leqq n - 1$. Moreover, it follows from 3.42 that $\beta_n(a, M; \mathfrak{f}) \geqq 1$. *We assert in fact that $\beta_n(a, M; \mathfrak{f}) = 1$.* For we can show that if $A(a)$ is a neighborhood of $a (a \subset L)$, and if Γ_n is a cycle mod $(M - A)$ there exists an $A_1(a) \subset A(a)$ and an $x \in \mathfrak{f}$ such that $\Gamma_n \sim x\Delta_n$ mod $(M - A_1)$. The proof of this is very much like that of 3.43 and we shall not insist on the details.

3.45. *Let $A(a)$ be an invariant neighborhood of a, $a \subset L$, and let Γ_h $(0 \leqq h \leqq n - 2)$ be a cycle* mod $(M - A)$ *of type $(\rho, M - A)$. There exists an invariant $A_1(a) \subset A(a)$, a cycle Γ_{h+1}* mod $(M - A_1)$ *of type $(\rho, M - A_1)$ and a cycle Γ_h' of type ρ (hence of type $(\rho, M - A_1)$ such that $\Gamma_{h+1}:\Gamma_h'$ and $\Gamma_h \simeq \Gamma_h'$* mod $(M - A_1)$. The proof is in principle the same as that of 3.24. The essential difference is that now we make use of 3.43 and 3.44 instead of III. There is no difficulty in making the necessary modifications and we may therefore omit details.

3.46. *Let $A(a)$ be an invariant neighborhood of a, $a \subset L$. There exists an invariant $A_1(a) \subset A(a)$ such that if Γ_h $(0 \leqq h \leqq n)$ is a cycle* mod $(M - A)$ *of type $(\rho_h, M - A)$, then $\Gamma_h \simeq x\Delta_h$* mod $(M - A_1 + L)$, $x \in \mathfrak{f}$. The proof makes use of 3.45 and is in principle the same as that of 3.40.

3.47. *Let K be an invariant set and let γ_h $(0 \leqq h \leqq n - 1)$ be a cycle* mod *KL in L. If X_{h+1} is a cycle* mod $(K + L)$ *such that $\phi X_{h+1} = \gamma_h$* mod K (see 1.1) *and if $\rho = \bar{\sigma}$ or if $\mathfrak{f} = \mathfrak{p}$ and $\rho = \sigma$, then $\Gamma_{h+1} = \{\rho X_{h+1}(\mathfrak{U})\}$ is a cycle* mod K *of type ρ.*

PROOF. $\phi\rho X_{h+1}(\mathfrak{U}) = \rho\phi X_{h+1}(\mathfrak{U}) = \rho\gamma_h(\mathfrak{U})$ mod K. Since $\gamma^1(\mathfrak{U}) = \gamma(\mathfrak{U})$ (3.5), we have $\rho\gamma_h(\mathfrak{U}) = 0$ mod K so that $\rho X_{h+1}(\mathfrak{U})$ is a cycle mod K. Let $\mathfrak{U} \subset \mathfrak{B}$, $\pi = \pi(\mathfrak{U}, \mathfrak{B})$. We have $\pi X(\mathfrak{B}) \sim X(\mathfrak{U})$ mod $(K + L)$ which implies a relation of the form

$$\phi Y_{h+2}(\mathfrak{U}) = \pi X(\mathfrak{B}) - X(\mathfrak{U}) + H(\mathfrak{U})$$

where $H(\mathfrak{U}) \subset K + L$. From 3.33, $\rho H(\mathfrak{U}) \subset K$; hence $\phi\rho Y(\mathfrak{U}) = \pi\rho X(\mathfrak{B}) - \rho X(\mathfrak{U})$ mod K, that is, $\pi\Gamma(\mathfrak{B}) \simeq \Gamma(\mathfrak{U})$ mod K, which establishes the proof.

3.48. We return now to a consideration of the cycles $\Delta_n, \cdots, \Delta_0$ of 3.32. We observe first that Δ_n is not $\simeq 0$ mod L, otherwise we would have $\Delta_n = 0$ mod L which is impossible (3.37). On the other hand $\Delta_0 \simeq 0$ mod L. For, since $B_0(M; \mathfrak{f}) = 0$, each covering $\mathfrak{U}$ is a connected complex ([1], p. 170). Hence if E is an arbitrary vertex of $\mathfrak{U}$, then E, regarded as a $(0, \mathfrak{U})$-cycle mod L is ~ 0 mod L. Hence we have $\chi_0(\mathfrak{U}) \sim 0$ mod L and therefore (3.16), $\Delta_0(\mathfrak{U}) = \rho_0\chi_0(\mathfrak{U}) \simeq 0$ mod L for every $\mathfrak{U}$.

In the sequence $\Delta_n, \Delta_{n-1}, \cdots, \Delta_0$, let Δ_r be the first cycle which is $\simeq 0$ mod L. The value of r depends of course on $\mathfrak{f}$ and we shall on occasion write $r = r(\mathfrak{f})$. In any case it follows from the preceding paragraph that $0 \leqq r \leqq n - 1$, and from 3.15, that $\Delta_{r-1} \simeq \Delta_{r-2} \simeq \cdots \simeq \Delta_0 \simeq 0$ mod L.

3.49. *Let* $\mathfrak{f} = \mathfrak{p}$ *and let* A *be an invariant neighborhood of* a, $a \subset L$. *For* $r < h \leqq n$, Δ_h *is not* $\simeq 0$ mod $(M - A + L)$.

Proof. Suppose on the contrary that for each $\mathfrak{U}$ there exists an $X_{h+1}(\mathfrak{U})$ such that $\phi\rho X_{h+1}(\mathfrak{U}) = \Delta_h(\mathfrak{U})$ mod $(M - A + L)$.

Let $D(\mathfrak{U}) = A \cap \Delta_h(\mathfrak{U})$, $C(\mathfrak{U}) = \Delta_h(\mathfrak{U}) - D(\mathfrak{U})$. Then $\phi\rho X_{h+1}(\mathfrak{U}) = D(\mathfrak{U})$ mod $(M - A + L)$, or

$$\phi\rho X_{h+1}(\mathfrak{U}) = D(\mathfrak{U}) + H(\mathfrak{U}) \tag{1}$$

say, where $H(\mathfrak{U}) \subset M - A + L \subset M - A$. Since $\phi D(\mathfrak{U}) = -\phi H(\mathfrak{U})$ and $\phi D(\mathfrak{U}) = -\phi C(\mathfrak{U})$, the chain $C(\mathfrak{U}) - H(\mathfrak{U})$ is an absolute cycle in $M - A$. Since Δ_h is of type (ρ_h, L), hence of type ρ_h (on account of 3.35), and since A is invariant, it is clear that $C(\mathfrak{U})$ and $D(\mathfrak{U})$ are also of type ρ_h. Therefore $H(\mathfrak{U}) = \phi\rho_h X_{h+1}(\mathfrak{U}) - D(\mathfrak{U})$ is also of type ρ_h and therefore so is $C(\mathfrak{U}) - H(\mathfrak{U})$, say $C(\mathfrak{U}) - H(\mathfrak{U}) = \rho_h Y_h(\mathfrak{U})$. Now the chains and cycles with which we are dealing are defined for each $\mathfrak{U}$. Therefore we may assume that the cycle $\rho_h Y_h(\mathfrak{U})$ is essential in $M - A$, for if it is not, the chains $C(\mathfrak{U})$, $D(\mathfrak{U})$, $\Delta(\mathfrak{U})$, etc. can be replaced by the projections into $\mathfrak{U}$ of their coefficients in a suitable normal refinement. It follows from III that we may write $\rho_h Y_h(\mathfrak{U}) = \phi Y_{h+1}(\mathfrak{U})$, $Y_{h+1}(\mathfrak{U}) \subset M - A_1$, where $a \subset A_1 = A_1^1 \subset A$. $\rho_{h+1} Y_{h+1}(\mathfrak{U})$ is a cycle in $M - A_1$, and since it also may be assumed to be essential, we have $\rho_{h+1} Y_{h+1}(\mathfrak{U}) = \phi Y_{h+2}(\mathfrak{U})$, $Y_{h+2}(\mathfrak{U}) \subset M - A_2$. Continuing in this manner we obtain a cycle $\rho_n Y_n(\mathfrak{U})$ in $M - A_{n-h}$ and $\rho_n Y_n(\mathfrak{U}) \sim 0$, hence $\rho_n Y_n(\mathfrak{U}) = 0$ mod L, hence $\simeq 0$ mod L. Since

$$\rho_n Y_n(\mathfrak{U}) : \rho_{n-1} Y_{n-1}(\mathfrak{U}) : \cdots : \rho_h Y_h(\mathfrak{U}) = C(\mathfrak{U}) - H(\mathfrak{U}),$$

we have (3.15)

$$C(\mathfrak{U}) - H(\mathfrak{U}) \simeq 0 \text{ mod } L. \tag{2}$$

But (1) implies that $D(\mathfrak{U}) + H(\mathfrak{U}) \simeq 0$ mod L. On adding this to (2), we have $C(\mathfrak{U}) + D(\mathfrak{U}) = \Delta_h(\mathfrak{U}) \simeq 0$ mod L for every $\mathfrak{U}$, and this is impossible.

Let A be an invariant neighborhood of a, $a \subset L$. We have of course $\Delta_h \simeq 0$ mod $(M - A + L)$ when $0 \leqq h \leqq r(\mathfrak{p})$. The preceding result however asserts

that Δ_h is not $\simeq 0 \bmod (M - A + L)$ when $r(\mathfrak{p}) < h \leqq n$. We may express this state of affairs by saying that the *local* value of $r(\mathfrak{p})$ at points of L is constant and equal to its value *in the large*. We do not know at present whether or not this holds for $\mathfrak{f} = \mathfrak{r}$, although a modification of the proof shows that the local value of $r(\mathfrak{r})$ can not exceed its value in the large by more than one.

3.50. *Let $A(a)$ be an invariant neighborhood of a, $a \subset L$. If $\mathfrak{f} = \mathfrak{p}$ and if M is HCL^n* (i.e. *locally connected with respect to homologies in the dimensions* $0, \cdots, n$) *there exists an invariant $A_1(a)$ such that every cycle of type ρ in A_1 is* $\simeq 0 \bmod L$ *in A*. The proof need not be given in detail; it resembles the latter part of the proof of 3.49 in that it is based on a building up process followed by a splitting down process.

3.51. *If $r(\mathfrak{f}) \leqq n - 2$* ($\mathfrak{f} = \mathfrak{p}$ or $\mathfrak{r}$) *there exists a cycle X_{r+1}* mod L *which is not* $\sim 0 \bmod L$. *If $\mathfrak{f} = \mathfrak{p}$, X_{r+1} is not* $\sim 0 \bmod (M - A + L)$ *where A is an arbitrary neighborhood of a, $a \subset L$.*

Proof. Since $\Delta_r \simeq 0 \bmod L$, there exists a relation of the form $\phi\rho_r H_{r+1}\mathfrak{U} = \Delta_r(\mathfrak{U}) \bmod L$ for every $\mathfrak{U}$. Since $\phi\chi_{r+1}(\mathfrak{U}) = \Delta_{r+1}(\mathfrak{U})$, the chain $X'_{r+1}(\mathfrak{U}) = \rho_r H_{r+1}(\mathfrak{U}) - \chi_{r+1}(\mathfrak{U})$ is a cycle mod L. Let $\mathfrak{U}_1$ be a normal refinement of $\mathfrak{U}$ relative to cycles mod L and let $X_{r+1}(\mathfrak{U}) = \pi_1 X'(\mathfrak{U}_1)$, $\pi_1 = \pi_1(\mathfrak{U}_1, \mathfrak{U})$. Then since all $(r + 1)$-cycles in M are ~ 0, it follows from 1.6 that $X_{r+1} = \{X_{r+1}(\mathfrak{U})\}$ is a cycle mod L. We have

$$(1) \quad \rho_{r+1} X_{r+1}(\mathfrak{U}) = \pi_1 \bar{\rho}_r \rho_r H_{r+1}(\mathfrak{U}_1) - \pi_1 \rho_{r+1} \chi_{r+1}(\mathfrak{U}_1)$$
$$= \pi_2 \Delta_{r+1}(\mathfrak{U}_1) \simeq \Delta_{r+1}(\mathfrak{U}) \bmod L.$$

If $X_{r+1} \sim 0 \bmod L$, then by 3.16 $\rho_{r+1} \simeq 0 \bmod L$, that is $\Delta_{r+1} \simeq 0 \bmod L$ which is impossible. If $\mathfrak{f} = \mathfrak{p}$ and if $X_{r+1} \sim 0 \bmod (M - A + L)$ then also $X_{r+1} \sim 0 \bmod (M - A_1 + L)$ where $a \subset A_1 = A_1^1 \subset A$ and hence $\rho_{r+1} X_{r+1} \simeq 0 \bmod (M - A_1 + L)$ which by 3.49 is impossible.

3.52. *Let $\mathfrak{f} = \mathfrak{p}$ and let $A(a)$ be an invariant neighborhood of a, $a \subset L$. Let $X_h (0 \leqq h \leqq n)$ be a cycle* mod $(M - A + L)$ *such that ρX is a cycle* mod $(M - A)$, *of type* $(\rho, M - A)$. *If $\rho X_h \simeq 0 \bmod (M - A)$ then for each $\mathfrak{U}$ there exists a cycle $X_{0,h}(\mathfrak{U})$* mod $(M - A)$ *such that* $X_h(\mathfrak{U}) = X_{0,h}(\mathfrak{U}) \bmod (M - A + L)$.

Proof. The relation $\rho X_h \simeq 0 \bmod (M - A)$ implies a relation of the form $\phi\rho Y_{h+1}(\mathfrak{U}) = \rho X_h(\mathfrak{U}) \bmod (M - A)$ for every $\mathfrak{U}$. Let

$$(1) \qquad Z_h(\mathfrak{U}) = \phi Y_{h+1}(\mathfrak{U}) - X_h(\mathfrak{U}).$$

Then $\rho Z_h(\mathfrak{U}) = 0 \bmod (M - A)$, hence mod $(M - A + L)$ and therefore by 3.12, we may write $Z_h(\mathfrak{U}) = \bar{\rho} Z'(\mathfrak{U}) \bmod (M - A + L)$. Now since $X_h(\mathfrak{U})$ is a cycle mod $(M - A + L)$, (1) shows that the same is true of $Z_h(\mathfrak{U})$. Hence $\bar{\rho} Z'_h(\mathfrak{U})$ is a cycle mod $(M - A + L)$, hence it is also a cycle mod $(M - A)$ by 3.35. Let $X_{0,h}(\mathfrak{U}) = \phi Y_h(\mathfrak{U}) - \bar{\rho} Z'(\mathfrak{U})$. Then $X_{0,h}(\mathfrak{U})$ is a cycle mod $(M - A)$ and we have $X_{0,h}(\mathfrak{U}) = X_h(\mathfrak{U}) \bmod (M - A + L)$.

4. Homologies in L

4.1. We are now in a position to draw certain conclusions concerning the homology characters of L. It will be sufficient to examine the relative and absolute M-cycles in L (i.e. the cycles defined in terms of the coverings of M). For, as Čech has pointed out, the homology groups of L defined in terms of L-cycles—that is, intrinsically in terms of coverings of the space L (regarded as a subspace of M) are the same as the groups defined in terms of the M-cycles in L (see [1] p. 168).

4.2. *If $r(\mathfrak{f}) \leqq n - 2$ there exists an absolute r-cycle δ_r in L which is not ~ 0 in L. If $\mathfrak{f} = \mathfrak{p}$, δ_r is not ~ 0 mod $L(M - A)$ in L where A is an arbitrary neighborhood of a, $a \subset L$.*

Proof. Consider the cycle X_{r+1} of 3.51 and let $\delta_r = \phi X_{r+1}$ (see 1.2). If $\delta_r \sim 0$ in L, it follows from 1.3 (since every $(r + 1)$-cycle in M is ~ 0) that $X_{r+1} \sim 0$ mod L which is impossible. If $\mathfrak{f} = \mathfrak{p}$ and if $\delta_r \sim 0$ mod $(M - A)L$ in L, then since there exists an $A_1(a) \subset A(a)$ such that every $(r + 1)$-cycle mod $(M - A)$ is ~ 0 mod $(M - A_1)$ (3.43) we may again apply 1.3; we conclude that $X_{r+1} \sim 0$ mod $(M - A_1 + L)$ which is impossible.

4.3. *The second part of* 4.2 *holds for* $r = n - 1$.

Proof. We have $\Delta_{n-1}(\mathfrak{U}) = \phi\rho_n H_n(\mathfrak{U})$ mod L so that

$$Y_n(\mathfrak{U}) = \chi_n(\mathfrak{U}) - \rho_n H_n(\mathfrak{U})$$

is a cycle mod L. Let $\phi Y_n(\mathfrak{U}) = \delta_{n-1}(\mathfrak{U}) =$ a $\mathfrak{U}$-cycle in L. Since $\Lambda\rho_n H_n(\mathfrak{U}) = 0$, we have $\Lambda Y_n(\mathfrak{U}) = \Lambda\chi_n(\mathfrak{U})$ and $\phi\Lambda\chi_n(\mathfrak{U}) = \Lambda\delta_{n-1}(\mathfrak{U}) = \delta^*_{n-1}(\mathfrak{U}^*)$ (say) $=$ a cycle in L^*. But by 3.19 $\{\Lambda\chi_n(\mathfrak{U})\}$ is a cycle mod L^* in M^*; hence $\delta^*_{n-1} = \{\delta^*_{n-1}(\mathfrak{U}^*)\}$ is a cycle in L^* (1.2). But by 3.6, Λ induces a correspondence between cycles in L and cycles in L^* in such a manner that all homology relations are preserved. We conclude that $\delta_{n-1} = \{\delta_{n-1}(\mathfrak{U})\}$ is a cycle in L. Suppose $\delta_{n-1} \sim 0$ mod $L(M - A)$ in L, say $\delta_{n-1}(\mathfrak{U}) = \phi Z_n(\mathfrak{U})$ mod $L(M - A)$, $Z(\mathfrak{U}) \subset L$. Then $Y_n(\mathfrak{U}) - Z_n(\mathfrak{U})$ is a cycle mod $(M - A)$ and we may assume as usual that this cycle is essential. Therefore on account of 3.44 there exists an $A_1(a)$ which we may assume to be invariant, such that $A_1 \subset A$ and $Y_n(\mathfrak{U}) - Z_n(\mathfrak{U}) \sim x\Delta_n(\mathfrak{U})$ mod $(M - A_1)$, $x \in p$. Since $Z_n(\mathfrak{U}) \subset L$, we have $Y_n(\mathfrak{U}) \sim x\Delta_n(\mathfrak{U})$ mod $(M - A_1 + L)$. Hence $\rho_n Y_n(\mathfrak{U}) \simeq x\rho_n\Delta_n(\mathfrak{U})$ mod $(M - A_1 + L)$ by 3.16; but $\rho_n\Delta_n(\mathfrak{U}) = p\Delta_n(\mathfrak{U}) = 0$ (mod p), hence $\Delta_n(\mathfrak{U}) \simeq 0$ mod $(M - A_1 + L)$ for every $\mathfrak{U}$; hence by 3.35 $\Delta_n \simeq 0$ mod $(M - A_1)$ which is a contradiction to 3.42.

Corollaries of 4.2 and 4.3: $\beta_r(a, L; \mathfrak{p}) \geqq 1$. *Also (taking $A = M$): $B_r(L; \mathfrak{p}) \geqq 1$, and when $r \leqq n - 2$, $B_r(L; \mathfrak{r}) \geqq 1$.*

4.4. We shall make one more assumption concerning M:

VI. Let Γ_{n-1} be a cycle in $G \subset M$ and let a, b, c be distinct points in $M - G$. For at least two of the points, say a, b, there are neighborhoods $A(a)$, $B(b)$ such that $\Gamma_{n-1} \sim 0$ in $M - (A + B)$.

4.5. *$r(\mathfrak{p})$ can equal $n - 1$ only if $p = 2$.*

PROOF. Suppose $p > 2$ and $r(\mathfrak{p}) = n - 1$. By 4.3 there exists a cycle γ_{n-1} in L which is not ~ 0 in L. Let b be a point in $M - L$. The points b, b^1, b^2 are distinct. By VI there exist neighborhoods of two of the points, say $B(b)$ and $B_1(b^1)$ such that $\gamma_{n-1} \sim 0$ in $M - (B + B_1)$. On replacing B by a smaller neighborhood if necessary, we may assume that $B_1 = B^1$. Thus for each $\mathfrak{U}$ we have $\gamma_{n-1}(\mathfrak{U}) = \phi Y_n(\mathfrak{U})$, $Y_n(\mathfrak{U}) \subset M - (B + B^1) \subset M - B^1$. Obviously $Y_n^1(\mathfrak{U}) \subset M - (B^1 + B^2) \subset M - B^1$ and $\phi Y_n^1(\mathfrak{U}) = \gamma_{n-1}^1(\mathfrak{U}) = \gamma_{n-1}(\mathfrak{U})$. The absolute cycle $Y_n(\mathfrak{U}) - Y_n^1(\mathfrak{U})$ is therefore in $M - B^1$. We may by the usual argument assume that $Y_n(\mathfrak{U}) - Y_n^1(\mathfrak{U})$ is essential in $M - B^1$. Hence $Y_n(\mathfrak{U}) - Y_n^1(\mathfrak{U}) \sim 0$, hence $Y_n(\mathfrak{U}) = Y^1(\mathfrak{U})$ mod L, hence by (3.11) $Y_n(\mathfrak{U}) = \sigma Z_n(\mathfrak{U})$ (say) mod L. But in this case it follows from 3.35 that $Y_n(\mathfrak{U})$ is an absolute cycle, hence for each $\mathfrak{U}$, $\gamma_{n-1}(\mathfrak{U}) = \phi Y_n(\mathfrak{U}) = 0$ which is impossible.

4.6. *$M - L$ can have more than one component only if $r(\mathfrak{p}) = r(\mathfrak{r}) = n - 1$.*

PROOF. Suppose that $M - L$ has more than one component. Then (by 3.38) $p = 2$ and the components are say A, A^1. Moreover $\Delta_n^1 \sim -\Delta_n$. Now consider the chain ${}^{(3)}D(\mathfrak{U}_3)$ in the proof of 3.38. It was shown that

$$\bar{\sigma}^{(3)}D(\mathfrak{U}_3) \sim x\Delta_n(\mathfrak{U}) \tag{1}$$

and that $x \neq 0$. Now if $\mathfrak{f} = \mathfrak{p}$, $\bar{\sigma}^{(3)}D(\mathfrak{U}_3) = \sigma^{(3)}D(\mathfrak{U}_3) = \rho_n{}^{(3)}D(\mathfrak{U}_3)$; if $\mathfrak{f} = \mathfrak{r}$, it follows from the relation $\Delta_n^1 \sim -\Delta_n$ that $\rho_n = \bar{\sigma}$. Thus in either case (1) implies that

$$\rho_n({}^{(3)}D(\mathfrak{U}_3) - x\chi_n(\mathfrak{U}_3)) \simeq 0 \text{ mod } L.$$

Hence by 3.15

$$\phi({}^{(3)}D(\mathfrak{U}_3) - x\chi_n(\mathfrak{U}_3) \simeq 0 \text{ mod } L.$$

But $\phi^{(3)}D(\mathfrak{U}_3) = 0$ mod L and $\phi\chi_n(\mathfrak{U}_3) = \Delta_{n-1}(\mathfrak{U}_3)$. Thus we have $\Delta_{n-1}(\mathfrak{U}_3) \simeq 0$ mod L. Since every $\mathfrak{U}$ possesses a refinement of the sort $\mathfrak{U}_3$, we have $\Delta_{n-1} \simeq 0$ mod L and hence $r(\mathfrak{p}) = r(\mathfrak{r}) = n - 1$.

4.7. *Let $\mathfrak{f} = \mathfrak{p}$ and let $A(a)$ be an invariant neighborhood of a, $a \subset L$. Let γ_r be a cycle* mod $L(M - A)$ *in L. There exists an invariant $A_1(a) \subset A(a)$ and an $x \in \mathfrak{p}$ such that $\gamma_r \sim x\delta_r$* mod $L(M - A_1)$ *in L, where δ_r is the cycle defined in* 4.2 *and* 4.3. *If $A = M$, A_1 can be taken equal to M.*

PROOF. Assume first that $r \leqq n - 2$. Then by 3.43 there exists an $A_0(a) \subset A(a)$ such that $\gamma_r \sim 0$ mod $(M - A_0)$ and such that every $(r + 1)$-cycle mod $(M - A)$ is ~ 0 mod $(M - A_0)$. Therefore, by 1.5 with $K_1 = M - A_0$. there exists a cycle Y_{r+1} mod $(M - A_0 + L)$ and a cycle $\gamma_{0,r}$ mod $(M - A_0)L$ in L such that $\phi Y_{r+1} = \gamma_{0,r}$ mod $(M - A_0)$ and $\gamma_{0,r} \sim \gamma_r$ mod $(M - A_0)L$ in L. By 3.47, $\rho_{r+1}Y_{r+1}$ is a cycle mod $(M - A_0 + L)$ of type $(\rho_{r+1}, M - A_0 + L)$. Hence by 3.35, $\rho_{r+1}Y_{r+1}$ is a cycle mod $(M - A_0)$ of type

$(\rho_{r+1}, M - A_0)$. By 3.46, there exists an $A_{00}(a) \subset A_0(a)$ such that $\rho_{r+1}Y_{r+1} \simeq x\Delta_{r+1}$ mod $(M - A_{00})$; that is

$$(1) \qquad \rho_{r+1}(Y_{r+1}(\mathfrak{U}) - x\chi_{r+1}(\mathfrak{U}) \simeq 0 \bmod (M - A_{00})$$

for every $\mathfrak{U}$; this relation holds of course mod $(M - A_{00} + L)$. Let X_{r+1} be the cycle mod L of 3.51, such that $\phi X_{r+1} = \delta_r$. By (1), 3.51, we have $\rho_{r+1}X_{r+1}(\mathfrak{U}) \simeq \rho_{r+1}\chi_{r+1}(\mathfrak{U})$ mod L, hence mod $(M - A_{00} + L)$. Therefore (1) becomes

$$\rho_{r+1}(Y_{r+1}(\mathfrak{U}) - xX_{r+1}(\mathfrak{U})) \simeq 0 \bmod (M - A_{00} + L).$$

Therefore by 3.15, we have

$$\phi(Y_{r+1}(\mathfrak{U}) - xX_{r+1}(\mathfrak{U})) \simeq 0 \bmod (M - A_{00} + L).$$

By 3.35, this last relation holds mod $(M - A_{00})$ and $\phi(Y_{r+1}(\mathfrak{U}) - xX_{r+1}(\mathfrak{U}))$ is a cycle mod $(M - A_{00})$ of type $(\rho_{r+1}, M - A_{00})$. By the usual argument we may assume that $\phi(Y_{r+1}(\mathfrak{U}) - xX_{r+1}(\mathfrak{U}))$ is essential and therefore by 3.43 there exists an $A_1(a)$ which we may assume to be invariant, such that $A_1 \subset A_{00}$ and such that $\phi(Y_{r+1}(\mathfrak{U}) - xX_{r+1}(\mathfrak{U})) \sim 0$ mod $(M - A_1)$. But

$$\phi(X_{r+1} - xX_{r+1}) = \gamma_r - x\delta_r \bmod (M - A_1).$$

Hence by 1.4, $\gamma_r - x\delta_r \sim 0$ mod $(M - A_1)L$ in L. It will be seen that in case $A = M$, then A_0, A_{00}, and A_1 can all be taken equal to M; we must then appeal to V (3.33) rather that 3.43.

We have still to consider the case $r(\mathfrak{p}) = n - 1$. For simplicity let us take $A = M$ and thus deal with absolute cycles; the modifications necessary for relative cycles will be evident.

We have now $p = 2$ (4.5). Let us suppose that γ_{n-1} is not ~ 0 in L. Since $\gamma_{n-1} \sim 0$ (in M) we may write $\gamma_{n-1}(\mathfrak{U}) = \phi Y_n(\mathfrak{U})$. Now by 3.35, $\sigma Y_n(\mathfrak{U})$ is an absolute cycle and we may as usual assume that $\sigma Y_n(\mathfrak{U})$ is essential, hence we have $\sigma Y_n(\mathfrak{U}) \sim x\Delta_n(\mathfrak{U})$ where (since $p = 2$), $x = 0$ or 1. Suppose $x = 0$. Then we have $\sigma Y_n(\mathfrak{U}) \sim 0$, hence $\sigma Y_n(\mathfrak{U}) = 0$ mod L, hence by 3.11, $Y_n(\mathfrak{U}) = \bar{\sigma} Y_{0,n}(\mathfrak{U})$ mod L, and $\phi Y_n(\mathfrak{U}) = \bar{\sigma}\phi Y_{0,n}(\mathfrak{U})$ mod L. But by 3.33, $L \cap \bar{\sigma}\phi Y_{0,n}(\mathfrak{U}) = 0$, therefore since $\phi Y_n(\mathfrak{U}) = \gamma_{n-1}(\mathfrak{U}) \subset L$, we have $\phi Y_n(\mathfrak{U}) = \gamma_{n-1}(\mathfrak{U}) = 0$ for every $\mathfrak{U}$, which is impossible. We conclude that $x = 1$ so that $\sigma Y_n(\mathfrak{U}) \sim \Delta_n(\mathfrak{U})$. In the same manner, since δ_{n-1} is not ~ 0 in L, we have $\sigma X_n(\mathfrak{U}) \sim \Delta_n(\mathfrak{U})$. Therefore $\sigma(X_n(\mathfrak{U}) - Y_n(\mathfrak{U})) = 0$ mod L, so that $X_n(\mathfrak{U}) - Y_n(\mathfrak{U}) = \bar{\sigma} Z(\mathfrak{U})$ mod L (say), and hence as above,

$$\phi(X_n(\mathfrak{U}) - Y_n(\mathfrak{U})) = 0$$

—that is, $\delta_{n-1} - \gamma_{n-1} = 0$.

Corollary of 4.7 (with reference to the corollary of 4.3): $\beta_r(a, L; \mathfrak{p}) = B_r(L; \mathfrak{p}) = 1$. *In particular, if* $r(\mathfrak{p}) = 0$, L *has two components* (see 3.23 and [1], pp. 168–170).

4.8. *Let $\mathfrak{f} = \mathfrak{p}$ and let $A(a)$ be a neighborhood of a, $a \subset L$. Let γ_h be a cycle* mod $(M - A)L$ *in L. If $0 \leqq h \leqq n$ and $h \neq r$, there exists an $A_1(a) \subset A(a)$ (with $A_1 = M$ if $A = M$) such that $\gamma_h \sim 0$* mod $(M - A_1)L$ *in L.*

Proof. Suppose $h = n$. Then by 3.44 we have $\gamma_n \sim x\Delta_n$ mod $(M - A_1)$ where $a \subset A_1 \subset A$. If $A = M$, we may take $A_1 = M$ by V (3.30). Obviously we may assume that $A_1 = A_1^1$. The homology then implies that $\gamma_n \simeq x\Delta_n$ mod $(M - A_1 + L)$. Since $\gamma_n \subset L$, we have $x\Delta_n \simeq 0$ mod $(M - A_1 + L)$. Hence by 3.49 (or by 3.37 when $A_1 = M$), $x = 0$. Thus $\gamma_n \sim 0$ mod $(M - A_1)$. Hence we may write

$$\gamma_n(\mathfrak{U}) = \phi Y_{n+1}(\mathfrak{U}) + H_n(\mathfrak{U}) \tag{1}$$

where $H_n(\mathfrak{U}) \subset M - A_1$. Since every $\mathfrak{U}$-simplex not in L is of dimension $\leqq n$ (2.16) we have $Y_{n+1}(\mathfrak{U}) \subset L$. Hence, since $\gamma_n(\mathfrak{U}) \subset L$, it follows from (1) that $H_n(\mathfrak{U}) \subset L$. Therefore, since M is completely normal we may assume that $H_n(\mathfrak{U}) \subset (M - A_1)L$ (Cf. footnote 8 and the proof of 1.1) and hence (1) implies that $\gamma_n \sim 0$ mod $(M - A_1)L$ in L.

Suppose that $h \leqq n - 2$. Then by 1.5 there exists in A an $A_0(a)$ which we may assume to be invariant, a cycle $X_{0,h+1}$ mod $(M - A_0 + L)$ and a cycle $\gamma_{0,h}$ mod $(M - A_0)L$ in L such that $\phi X_{0,h+1} = \gamma_{0,h}$ mod $(M - A_0)$ in L and $\gamma_{0,h} \sim \gamma_h$ mod $(M - A_0)L$ in L. Now by 3.47 $\rho_{h+1}X_{0,h+1}$ is a cycle mod $(M - A_0)$ of type $(\rho_{h+1}, M - A_0)$ and is therefore $\simeq x\Delta_{h+1}$ mod $(M - A_{00} + L)$ where $a \subset A_{00} \subset A_{00}^1 \subset A_0$ (3.46). If $h < r$ we have $\rho_{h+1}X_{0,h+1} \simeq 0$ mod $(M - A_{00} + L)$. This same relation holds mod $(M - A_{00})$ by 3.35, and hence we have $X_{0,h+1}(\mathfrak{U}) = X'_{0,h+1}(\mathfrak{U})$ mod $(M - A_{00} + L)$ where $X'_{0,h+1}(\mathfrak{U})$ is a cycle mod $(M - A_{00})$ (3.52). Since $h + 1 \leqq n - 1$, and since $X'_{0,h+1}(\mathfrak{U})$ may be assumed to be essential mod $(M - A_{00})$, we have (by 3.43) $X'_{0,h+1}(\mathfrak{U}) \sim 0$ mod $(M - A_1)$ where $a \subset A_1 \subset A_{00}$, and hence $X_{0,h+1}(\mathfrak{U}) \sim 0$ $(M - A_1 + L)$ and therefore $\gamma_{0,h} \sim 0$ mod $(M - A_1)L$ in L by 1.4, hence $\gamma_0 \sim 0$ mod $(M - A_1)L$ in L.

Suppose next that $r < h \leqq n - 2$. The relation $\rho_{h+1}X_{h+1} \simeq x\Delta_{h+1}$ mod $(M - A_{00} + L)$ can be written

$$\rho_{h+1}(X_{0,h+1}(\mathfrak{U}) - x\chi_{h+1}(\mathfrak{U})) \simeq 0 \text{ mod } (M - A_{00} + L).$$

Hence by 3.15, $\phi X_{0,h+1}(\mathfrak{U}) - x\chi_{h+1}(\mathfrak{U}) \simeq 0$ mod $(M - A_{00} + L)$, that is

$$\gamma_{0,h}(\mathfrak{U}) - x\Delta_h(\mathfrak{U}) \simeq 0 \text{ mod } (M - A_{00} + L).$$

But $\gamma_{0,h}(\mathfrak{U}) = 0$ mod $(M - A_{00} + L)$ and since Δ_h is not $\simeq 0$ mod $(M - A_{00} + L)$ (3.49), we must have $x = 0$. Thus $\rho_{h+1}X_{0,h+1} \simeq 0$ mod $(M - A_{00} + L)$ and we continue the argument as before.

We have finally to consider the case $h = n - 1$. Since γ_{h-1} is a cycle mod $(M - A)$, there exists by 3.43, an $X_n(\mathfrak{U})$ such that $\phi X_n(\mathfrak{U}) = \gamma_{n-1}(\mathfrak{U})$ mod $M - A_0$ $(a \subset A_0 \subset A)$. We may obviously assume A_0 invariant. Hence $\rho_n X_n(\mathfrak{U})$ is a cycle mod $(M - A_0)$ and as we may assume it to be essential, we have $\rho_n X_n(\mathfrak{U}) \sim x\Delta_n(\mathfrak{U})$ mod $(M - A_1)$ $(a \subset A_1 = A_1^1 \subset A_0)$. Now the

argument used in the preceding paragraph shows that $x = 0$. Hence $\rho_n X_n(\mathfrak{U}) = 0 \bmod (M - A_1 + L)$ and hence $X_n(\mathfrak{U})$ is a cycle mod $(M - A_1 + L)$ of type $(\bar{\sigma}_n, M - A_1 + L)$. Hence by 3.35 $X_n(\mathfrak{U})$ is a cycle mod $(M - A_1)$. It follows that $\phi X_n(\mathfrak{U}) = 0 \bmod (M - A_1)$ that is, $\gamma_{n-1}(\mathfrak{U}) = 0 \bmod (M - A_1)$. Thus $\gamma_{n-1}(\mathfrak{U}) \subset M - A_1$. Since also $\gamma_{n-1}(\mathfrak{U}) \subset L$, it follows readily (by an argument which makes use of a refinement $\mathfrak{U}_1$ of $\mathfrak{U}$ such that chains which are simultaneously in $M - A_1$ and L are in $(M - A_1)L$; see footnote 8) that $\gamma_{n-1}(\mathfrak{U}) \sim 0 \bmod (M - A_1)L$ in L. This completes the proof of 4.8.

Corollary of 4.8: $B_h(L; \mathfrak{p}) = \beta_h(a, L; \mathfrak{p}) = 0$ $(1 \leqq h \leqq n, h \neq r, a \subset L)$. *In particular if $r(\mathfrak{p}) > 0$, L is connected.*

From the corollaries of 4.7 and 4.8 we have now the following:

Theorem. *With respect to the coefficient field $\mathfrak{p}$, the Betti numbers of L, both locally and in the large, are the same as those of an $r(\mathfrak{p})$-sphere.*[11]

4.9. Alexandroff [2] has introduced certain local invariants that differ slightly from the local Betti numbers β_h. These new invariants are called the Betti numbers of L *in* a $(a \subset L)$ and are denoted by $p_a^h(L; \mathfrak{f})$. Let γ_h be a cycle mod $(M - A)L$ in L. In the language of Alexandroff, γ_h is a cycle in L *through* a. If there exists an $A_1(a)$ such that $\gamma_h \sim 0 \bmod (M - A_1)L$ in L, then γ_h is homologous to zero *in* a. The numbers $p_a^h(L; \mathfrak{f})$ are the Betti numbers of the h-cycles *through* a relative to homologies *in* a (the coefficient field being $\mathfrak{f}$). It will be seen from 4.2, 4.3, 4.7, 4.8, that the Betti numbers in a have the same values as the local Betti numbers at a,—thus $p_a^r(L; \mathfrak{p}) = 1$, $p_a^h(L; \mathfrak{p}) = 0$ when $h \neq r(\mathfrak{p})$ $(0 \leqq h \leqq n)$. Since $\mathfrak{p}$ is a field, it follows that relative to $\mathfrak{p}$ the Betti group in an arbitrary point a of L vanishes for each dimension greater than $r(\mathfrak{p})$ but not for $r(\mathfrak{p})$.

4.10. We may remark finally, before considering further restrictions on M, that if M is HLC^n with respect to the coefficient field $\mathfrak{p}$, then so is L. This follows readily from 3.50 and we shall omit details of proof.

4.11. Let us consider the special case in which the space M is compact and metric. We can now assume that M is immersed in a Euclidean space of a finite number of dimensions and therefore certain theorems of Alexandroff are applicable. In particular it follows from the last sentence of 4.9 and from [2], theorem II′, §7, that $\dim_p L = r(\mathfrak{p})$, where $\dim_{\mathfrak{f}} = L$ means the dimension of L relative to the field $\mathfrak{f}$ ([7]). Alexandroff has shown that $\dim_{\mathfrak{f}} L \leqq \dim L$

[11] Since $\dim L \leqq \dim M = n$, the h-dimensional Betti numbers of L are all zero when $h > n$. (This follows immediately from the existence for L of complete families of coverings of order $\leqq n$.)

([7], p. 195). Hence dim $L \geqq r(\mathfrak{p})$. Moreover, $\dim_{\mathfrak{p}} L \leqq n - 1$ (since $r(\mathfrak{p}) \leqq n - 1$).

Suppose $r(\mathfrak{p}) = 0$. Then L consists of two points (so that now dim $L = r(\mathfrak{p}) = r(\mathfrak{r}) = 0$). For in any case L has two components say K_1 and K_2. Let a be a point of K_1. If K_1 contains points other than a, it is easy to see that every relative $(0, \mathfrak{U})$-cycle mod $(M - A)L$ $(a \subset A)$ in L is ~ 0 mod $(M - A_1)L$ in L if $A_1(a)$ is sufficiently small. But this situation is impossible since $\beta_0(a, L; \mathfrak{p}) = 1$. Therefore K_1 is a single point and so, of course, is K_2.

If dim $L = 0$, then again L consists of two points, for we have $0 \leqq r(\mathfrak{p}) = \dim_{\mathfrak{p}} L \leqq \dim L = 0$ so that $r(\mathfrak{p}) = 0$.

Suppose that dim $L = 1$. Then $r(\mathfrak{p}) = 1$. For in any case $r(\mathfrak{p}) \leqq \dim L = 1$ and if $r(\mathfrak{p}) = 0$, we would have dim $L = 0$ as we have just seen. Since L is now connected and since $\beta_1(a, L; \mathfrak{p}) = 1$ for every $a \subset L$, it follows from a theorem of Alexandroff[12] that *L is a simple closed curve* and hence that dim $L = r(\mathfrak{p}) = 1$ and $r(\mathfrak{r}) \leqq 1$.

4.12. *If M is compact metric and if dim $L = 0$ and n is odd, then $\Delta_n^1 \sim -\Delta_n$.*

Proof. Since dim $L = 0$, we have $r(\mathfrak{p}) = 0$ (4.11). Moreover L consists of two points and hence there obviously exists in L a cycle δ_0 relative to the field $\mathfrak{f}$ of rational numbers, such that δ_0 is not ~ 0 in L and such that if all coefficients are reduced mod p, (so that $\mathfrak{f}$ becomes $\mathfrak{p}$), δ_0 is still not ~ 0 in L. Let X_1 be a cycle mod L and $\delta_{0,0}$ a cycle in L such that $\phi X_1 = \delta_{0,0}$, $\delta_{0,0} \sim \delta_0$ in L (1, 5). Then $\bar{\sigma} X_1$ is a cycle of type $\bar{\sigma}$. Now suppose that $\Delta_n^1 \sim \Delta_n$. Then $\rho_n = \sigma$ (3.31) and since n is odd we have $\rho_1 = \sigma$. Hence $\bar{\sigma} X_1$ is of type $\bar{\rho}_1$ and by 3.41, $\bar{\sigma} X_1 \simeq 0$ mod L. This relation obviously subsists if we reduce the field $\mathfrak{f}$ modulo p. Hence from 3.52 there exists an absolute cycle X_0' mod L (relative to $\mathfrak{p}$) such that $X_0 \sim X_0'$ mod L. From 1.2, $\delta_0' = \phi X_0'$ is a cycle in L and from 1.4 $\delta_0' \sim \delta_0$ in L. But since $\phi X_0' = 0$, we have $\delta_0 \sim \delta_{0,0} \sim 0$ in L which is impossible.

4.13. Suppose that M is locally euclidean. Then since L is nowhere dense in M, we have dim $L \leqq n$ and consequently *if $r(\mathfrak{p}) = n - 1$, we have dim $L = n - 1$*, since $r(\mathfrak{p}) \leqq \dim L$.

4.14. It is obvious that the assumptions I–VI hold when M is an n-sphere H_n. Hence all the results up to this point hold for H_n. In addition, *we have for*

[12] Corollary 1 of the theorem of p. 12 [8] asserts that in order that a one-dimensional continuum L be a simple closed curve, it is necessary and sufficient that the one-dimensional Betti number $p^1(a, L)$ (relative to the field of rational numbers) be one "around" each point a of L. For every field f, the numbers $p^h(a, L)$ relative to $\mathfrak{f}$ are equal to the local Betti numbers $\beta_h(a, L; \mathfrak{f})$. On several occasions in [8] it is pointed out that the results obtained hold equally well when the coefficient field is (say) $\mathfrak{p}$, and although it is not clear whether or not this is implied for the theorem just quoted, an examination of the proof shows readily that this theorem also, holds relative to $\mathfrak{p}$.

$M = H_n$ *the relation* $r(\mathfrak{r}) \leqq r(\mathfrak{p})$. For it is easy to show that $\Delta_n^{(\mathfrak{p})}$ (i.e. the cycle Δ_n relative to the field $\mathfrak{p}$) as well as the cycles $\Delta_{n-1}^{(\mathfrak{p})}$, $\Delta_{n-2}^{(\mathfrak{p})}$, etc. are obtainable from $\Delta_n^{(\mathfrak{r})}$, $\Delta_{n-1}^{(\mathfrak{r})}$ etc. simply by reducing all coefficients modulo p. Obviously the relation $\Delta_h^{(\mathfrak{r})} \simeq 0$ mod L implies $\Delta_h^{(\mathfrak{p})} \simeq 0$ mod L (mod p). Hence $r(\mathfrak{p}) \geqq r(\mathfrak{r})$.

4.15. Suppose that M is a 3-sphere. If T preserves orientation then L is a simple closed curve. For we now have $\Delta_n^1 \sim \Delta_n$ both for $\mathfrak{f} = \mathfrak{p}$ and $\mathfrak{f} = \mathfrak{r}$. Hence from 4.12, $r(\mathfrak{p}) \neq 0$. If $r(\mathfrak{p}) = 2$, then $B_2(L; \mathfrak{p}) = 2$ and hence by the Pontrjagin duality theorem [9] $M - L$ has two components. Hence by 3.37 $\Delta_n^1 \sim -\Delta_n$ which is impossible. We conclude that $r(\mathfrak{p}) \neq 2$, hence $r(\mathfrak{p}) = 1$. Now in a 3-sphere, the dimension of every closed subset equals its dimension modulo p ([7], page 209). Hence we have dim $L = \dim_{\mathfrak{p}} L = 1$, and therefore (4.11), L is a simple closed curve. We have thus the following

THEOREM. *If T is an orientation preserving transformation of a 3-sphere into itself and if the period to T is a prime number, then either T has no invariant points or else its invariant points constitute a simple closed curve.*

It seems probable that this theorem holds even if p is not a prime. Suppose for example that p contains the prime number p_1 as a factor, say $p = p_1q$. Then T^q is of period p_1 and hence, if T has any fixed points, they constitute a simple closed curve j. Now if T has any fixed points they constitute a closed subset j_0 of j and it seems improbable that j_0 can be a proper subset. For example if j contains an isolated point a, various reasons seem to indicate that T could not preserve orientation. We shall not, however, investigate this question further at present.

4.16. Suppose again that M is a 3-sphere, that p is a prime and that T reserves orientation. It is well known that T leaves fixed at least one point. From 3.31 we see that $p = 2$, $\rho_3 = \bar{\sigma}$. Suppose $r(\mathfrak{p}) = 1$. Let γ_1 be a cycle with coefficients in the field of rational numbers and such that γ_1 is not ~ 0 in L. Since L is now a simple closed curve, it is easy to see that γ_1 can be chosen in such a way that it will not be ~ 0 in L if all coefficients are reduced modulo 2. Let X_2 be a cycle mod L such that $\phi X_2 \sim \gamma_1$ in L (see 1.2 and 1.5). $\bar{\sigma}X_2$ is an absolute cycle and since $\bar{\sigma} = \bar{\rho}_2$, it follows from 3.41 that $\bar{\sigma}X_2 \simeq 0$. This relation subsists if all coefficients are reduced mod 2. Since $\mathfrak{f}$ now becomes $\mathfrak{p}$, we may apply 3.52 and infer that $\gamma_1 \sim 0$ in L (mod 2) precisely as in the case of δ_0 in the proof of 4.12. This contradiction proves that $r(\mathfrak{p}) = 1$ or 3. We have then the theorem that *if T is an orientation-reversing transformation of a 3-sphere into itself and if its period p is prime, then $p = 2$ and either there are two fixed points or else the set of fixed points is of dimension 2 and has the mod 2 Betti numbers of a 2-sphere both locally and in the large.*

Final Remarks

1. Suppose that p is not a prime. Let $p_1, \cdots, p_k$ be the prime factors of p and let $q_i = p/p_i$. If it is known that the set L of fixed points of T is identical with the set of fixed points of every power of T (a condition which is not in general satisfied as simple examples show) we can study L by considering the transformations $T^{q_1}, \cdots, T^{q_k}$. Since T^{q_i} is of period p_i it follows, for example, that if assumptions I–VI are satisfied, the Betti numbers of L, modulo q_i $(i = 1, \cdots, h)$ are the same as those of an r-sphere where $0 \leqq r \leqq n - 1$. There are various reasons why the methods which we have developed are not applicable directly to T. For example if a is an arbitrary point of M, it is no longer true that σa is a single point or else consists of p distinct points, and this condition played an essential part in the construction of special coverings. It seems certain, however, that the modifications which are necessary to make the whole development apply for arbitrary p, will not be difficult to carry out.

2. The homology characters of M^* modulo L^* can be computed by the aid of 3.19 and 3.20. It turns out readily that

$$\begin{aligned} B_h(M^*; L^*; \mathfrak{p}) &= 1 \text{ when } h = r + 1, \cdots, n \\ &= 0 \text{ when } h = 0, \cdots, r. \\ B_h(M^*; L^*; \mathfrak{r}) &= 1 \text{ when } h \text{ is of same parity as } n \text{ and } r < h \leqq n \\ &= 0 \text{ for all other values of } h. \end{aligned}$$

3. Although the integer $r(\mathfrak{p})$ depends in its definition on the manner in which the chains $\chi_n, \chi_{n-1}, \cdots$ are chosen, it is nevertheless topologically definite since it equals the dimension of the sphere whose Betti numbers mod p are the same as those of L. We have not shown however that $r(\mathfrak{r})$ is topologically definite. There seems to be some reason for thinking that under suitable restrictions (e.g. M = a hypersphere) if not in the most general case, $r(\mathfrak{p})$ must equal $r(\mathfrak{r})$ and both numbers must be of the same parity as n except when $p = 2$. We must however leave these questions open as well as the question of determining completely the Betti numbers of L other than those modulo p.

1. E. Čech, Théorie générale de l'homologie dans un espace quelconque, Fund. Math. 18 (1932), pp. 149–183.
2. E. Čech, Sur la dimension des espaces parfaitement normaux, Bull. intern. de l'acad. des Sc. de Bohème, 1932, pp. 38–55.
3. E. Čech, Théorie générale des variétés et de leurs théorèmes de dualité, Annals of Math. 34 (1933), pp. 621–730.
4. E. Čech, Sur les nombres de Betti locaux, Annals of Math. 35 (1934), pp. 678–701.
5. E. Čech, Sur la connexité locale d'ordre supérieur, Compositio Math. 2 (1935), pp. 1–26.
6. K. Menger, Dimensionstheorie (1928).
7. P. Alexandroff, Dimensionstheorie, Math. Annalen 106 (1932), pp. 161–238.
8. P. Alexandroff, On local properties of closed sets, Annals of Math. 36 (1935), pp. 1–35.

9. L. Pontrjagin, The general topological theorem of duality for closed sets, Annals of Math. 35 (1934), pp. 904–914.
10. M. H. A. Newman, A theorem on periodic transformations of spaces, Quart. Journ. 2 (1931) pp. 1–8.
11. P. Urysohn, Über die Mächtigkeit zuzammenhängander Mengen, Math. Annalen 94 (1925), pp. 262–295.
12. L. E. J. Brouwer, Über die periodischen Transformationen der Kugel, Math. Annalen 80 (1919), pp. 39–41.
13. B. Kérékjártó, Über die periodischen Transformationen der Kreisscheibe und Kugelflache, Math. Annalen, 80 (1919), pp. 36–38.
14. P. A. Smith, Transformations of period two, Rec. Math. Moscou, N. s. 1 (1936), pp. 761–764.

Barnard College,
Columbia University.

Annals of Mathematics
Vol. 39, No. 1, January, 1938

ON PROJECTIVE NORMAL COÖRDINATES

By Shiing-shen Chern*

(Received August 11, 1937)

1. **Introduction.** The normal coördinates in a projectively connected space were first introduced by O. Veblen and J. M. Thomas.[1] In a purely geometrical way E. Cartan has also defined a system of projective normal coördinates.[2] In this paper we shall give an analytic characterization of the normal coördinates of Cartan, from which it results that the two definitions are not the same. Further, some necessary conditions for the spaces for which the two definitions agree are obtained. I want to express here my hearty thanks to Professor Cartan for his valuable suggestions and criticisms.

2. **Analytic Characterization of the Normal Coördinates of Cartan.** Consider a space of n dimensions with the coördinates $u^1, \cdots, u^n$. According to Cartan a projective connection is defined in the space if there is given at each point A a projective reference ("repère projectif") $AA_1 \cdots A_n$ and a law of infinitesimal displacement between the references

$$(1)\qquad \begin{cases} dA \;= \qquad\quad \omega^1 A_1 + \cdots + \omega^n A_n, \\ dA_1 = \omega_1^0 A + \omega_1^1 A_1 + \cdots + \omega_1^n A_n, \\ \cdots\cdots \\ dA_n = \omega_n^0 A + \omega_n^1 A_1 + \cdots + \omega_n^n A_n, \end{cases}$$

where $\omega^i, \omega_i^0, \omega_i^j$ are differential forms in u^i, thus[3]

$$(2)\qquad \begin{cases} \omega^i = \Pi_k^i du^k, \qquad \omega_i^0 = \Pi_{ik}^0 du^k, \\ \omega_i^j = \Pi_{ik}^j du^k. \end{cases}$$

It is possible to change the projective reference $AA_1 \cdots A_n$ about its origin A without changing the projective connection. For a given system of coördinates u^i there is at each point a uniquely determined reference, called the natural reference ("repère naturel"), for which we have

$$(3)\qquad \Pi_k^i = \delta_k^i, \qquad \Pi_{ik}^i = 0.$$

* Research fellow of the China Foundation for the Promotion of Culture and Education.

[1] O. Veblen and J. M. Thomas, "Projective normal coördinates for the geometry of paths," Proc. N. A. S., vol. 11 (1925), pp. 204–7. See also T. Y. Thomas, Differential Invariants of Generalized Spaces, pp. 91–96.

[2] E. Cartan, Leçons sur la théorie des espaces à connexion projective, Paris 1936, pp. 220–4. This book will be cited as "Cartan, Leçons."

[3] For all details the reader is referred to Cartan, Leçons, Part 2.

For the sake of avoiding ambiguity we shall denote by Γ^j_{ik} the components Π^j_{ik} when the space is referred to the system of natural references. It may be verified that the components Γ^j_{ik} are the components of a projective connection in the sense of T. Y. Thomas.[4] This would mean that under a transformation of coördinates

$$(4) \qquad u^i = u^i(\bar{u}^1, \cdots, \bar{u}^n), \qquad \Delta = \left| \frac{\partial u^i}{\partial \bar{u}^k} \right| \neq 0,$$

they are transformed according to the equations

$$(5) \quad \bar{\Gamma}^l_{ik} \frac{\partial u^j}{\partial \bar{u}^l} = \frac{\partial^2 u^j}{\partial \bar{u}^i \partial \bar{u}^k} + \Gamma^j_{lm} \frac{\partial u^l}{\partial \bar{u}^i} \frac{\partial u^m}{\partial \bar{u}^k} - \frac{1}{n+1} \frac{\partial u^j}{\partial \bar{u}^k} \frac{\partial \log \Delta}{\partial \bar{u}^i} - \frac{1}{n+1} \frac{\partial u^j}{\partial \bar{u}^i} \frac{\partial \log \Delta}{\partial \bar{u}^k},$$

which play a fundamental part in the theory of Thomas.

The coördinates u^i are said to be normal coördinates in the sense of Veblen and Thomas[5] if the following relations are satisfied

$$(6) \qquad \Gamma^j_{ik} u^i u^k = 0.$$

The coördinates u^i remain normal when they undergo the transformation

$$(7) \qquad u^i \to \frac{a^i_k u^k}{1 + a^0_k u^k}, \qquad | a^i_k | \neq 0,$$

where the a^0_k, a^i_k are arbitrary constants.

The definition of Cartan's normal coördinates is purely geometrical. The coördinates u^i are said to be normal in Cartan's sense at the point $O(u^i = 0)$ if the geodesics through O are given by the equations

$$(8) \qquad u^i = a^i t$$

and if by developing the geodesics through O in the tangent space associated at O and by assigning the corresponding points the same coördinates u^i the coördinates u^i form a system of projective coördinates in the tangent space at O. The existence of Cartan's normal coördinates and the fact that they remain normal under the transformations (7) are geometrically evident. Analytically, in order that the coördinates u^i be normal it is necessary and sufficient that along the curves $u^i = a^i t$ (a^i being arbitrary constants) the projective acceleration of the geometrical point A be zero, i.e., there exists a factor $\lambda(u^1, \cdots, u^n) \neq 0$ such that, for $u^i = a^i t$,

$$(9) \qquad \frac{d^2(\lambda A)}{dt^2} = 0.$$

---

[4] T. Y. Thomas, "On the projective and equi-projective geometries of paths," Proc. N. A. S. vol. 11 (1925), pp. 199-203.

[5] Veblen and Thomas, loc. cit.

But we have, on referring to the system of natural references.

$$\frac{dA}{dt} = \frac{du^i}{dt} A_i,$$

$$\frac{d^2A}{dt^2} = \Pi^0_{ij} \frac{du^i}{dt} \frac{du^j}{dt} A + \left(\frac{d^2u^i}{dt^2} + \Gamma^i_{jk} \frac{du^j}{dt} \frac{du^k}{dt}\right) A_i.$$

Hence the condition (9) is equivalent to the conditions

$$\frac{\partial^2 \lambda}{\partial u^i \partial u^j} a^i a^j + \lambda\, \Pi^0_{ij} a^i a^j = 0,$$

$$2a^i a^k \frac{\partial \lambda}{\partial u^k} + \lambda\, \Gamma^i_{jk} a^j a^k = 0.$$

As this is true for all values of a^i, we get

$$(10)\qquad \begin{cases} \dfrac{\partial^2 \lambda}{\partial u^i \partial u^j} u^i u^j + \lambda \Pi^0_{ij} u^i u^j = 0, \\[2ex] 2 \dfrac{\partial \lambda}{\partial u^k} u^i u^k + \lambda \Gamma^i_{jk} u^j u^k = 0. \end{cases}$$

The necessary and sufficient conditions such that a solution $\lambda(u^1, \cdots, u^n) \neq 0$ of the system (10) exists, which is holomorphic at the origin, are

$$(11)\qquad \begin{cases} \Gamma^i_{jk} u^j u^k - 2Fu^i = 0, \\[2ex] \Pi^0_{ij} u^i u^j - \dfrac{\partial F}{\partial u^i} u^i + F^2 + F = 0, \end{cases}$$

where F is the common ratio of $\Gamma^i_{jk} u^j u^k : 2u^i$. Consequently, the conditions (11) are the necessary and sufficient conditions such that the coördinates u^i be normal in Cartan's sense.

3. **Question on the Coincidence of the Two Kinds of Normal Coördinates.** From the analytic definitions of the two kinds of normal coördinates it is evident that they are in general not the same. The question naturally arises of investigating those projectively connected spaces for which the two definitions agree. For a given point a necessary and sufficient condition that the normal coördinates u^i of Veblen and Thomas be also normal in the sense of Cartan is that they satisfy the equation $\Pi^0_{ik} u^i u^k = 0$.

A more difficult but also more interesting problem is to characterize intrinsically the projectively connected spaces for which the two kinds of normal coördinates always coincide. For this purpose we shall follow the manner by which Cartan studies his normal coördinates. Let O be a fixed point and R_0 a reference with the origin O. Suppose u^i be a system of normal coördinates with the origin $O(u^i = 0)$. In a sufficiently small region about O in which every point can be reached by one and only one geodesic through O, we get, by

displacing R_0 along all the geodesics $u^i = a^i t$, a uniquely determined reference at each point. Let

$$\omega^i = \Pi^i_k \, du^k, \qquad \omega^0_i = \Pi^0_{ik} \, du^k, \qquad \omega^j_i = \Pi^j_{ik} \, du^k$$

be the components of the projective connection with respect to this system of references. If we put $u^i = a^i t$ and vary t only, the reference undergoes a translation. That is to say, by putting $da^i = 0$ in

$$\begin{aligned} \omega^i &= \Pi^i_k \, t \, da^k + \Pi^i_k \, a^k \, dt, \\ \omega^0_i &= \Pi^0_{ik} \, t \, dx^k + \Pi^0_{ik} \, a^k \, dt, \\ \omega^j_i &= \Pi^j_{ik} \, t \, da^k + \Pi^j_{ik} \, a^k \, dt, \end{aligned}$$

we have

$$\omega^i = a^i \, dt, \qquad \omega^0_i = \omega^j_i = 0.$$

We get thus the conditions

$$\Pi^i_k \, a^k = a^i, \qquad \Pi^0_{ik} \, a^k = 0, \qquad \Pi^j_{ik} \, a^k = 0.$$

Since this holds for all values of a^i, we have

$$\Pi^i_k \, u^k = u^i, \qquad \Pi^0_{ik} \, u^k = 0, \qquad \Pi^j_{ik} \, u^k = 0. \tag{12}$$

The relations (12) characterize the space to be referred to a system of normal coördinates (in Cartan's sense) and to a system of references as defined above. If we define the quantities b^i_k by

$$b^i_k \, \Pi^k_j = \delta^i_j, \tag{13}$$

the components of Thomas are given by

$$\begin{aligned} \Gamma^j_{ik} = b^j_l \frac{\partial \Pi^l_i}{\partial u^k} + \Pi^h_i b^j_l \Pi^l_{hk} - \frac{1}{n+1} \delta^j_i \left(b^h_l \frac{\partial \Pi^l_h}{\partial u^k} + \Pi^h_{hk} \right) \\ - \frac{1}{n+1} \delta^j_k \left(b^h_l \frac{\partial \Pi^l_h}{\partial u^i} + \Pi^h_{hi} \right). \end{aligned} \tag{14}$$

The coördinates u^i are also normal in the sense of Veblen and Thomas when and only when

$$\Gamma^j_{ik} \, u^i u^k = 0,$$

i.e., when and only when,

$$\frac{\partial \, | \Pi^i_j |}{\partial u^k} \, u^k = 0,$$

or

$$| \Pi^i_j | = \text{constant}.$$

But $|\Pi^i_j| = 1$ at 0. Therefore the necessary and sufficient condition that the coördinates u^i be normal in both senses is that

$$[\omega^1 \cdots \omega^n] = [du^1 \cdots du^n]. \tag{15}$$

In order to give the condition (15) another form let us recall how the projective connection can be found when we know the values of the tensor of curvature and torsion and of its successive covariant derivatives

$$(R^{\beta}_{\alpha kh})_0,\ (R^{\beta}_{\alpha kh|l})_0,\ (R^{\beta}_{\alpha kh|lm})_0,\ \cdots \begin{pmatrix}\alpha, \beta = 0, 1, \cdots, n \\ R^0_{0kh} = 0\end{pmatrix}$$

at the point O. Let δ be an operation such that $\delta t = 0$, the δa^i being arbitrary. Put

$$\omega^i(\delta) = \bar{\omega}^i, \qquad \omega^j_i(\delta) = \bar{\omega}^j_i, \qquad \omega^0_i(\delta) = \bar{\omega}^0_i.$$

The system of differential equations[6]

$$\left\{\begin{aligned} \frac{\partial \bar{\omega}^i}{\partial t} &= \delta a^i + a^k \bar{\omega}^i_k - \{(R^i_{0kh})_0 + (R^i_{0kh|l})_0\, a^l t + \cdots\}\, a^k \bar{\omega}^h, \\ \frac{\partial \bar{\omega}^j_i}{\partial t} &= -a^j \bar{\omega}^0_i - \delta^j_i a^k \bar{\omega}^0_k - \{(R^j_{ikh})_0 + (R^j_{ikh|l})_0\, a^l t + \cdots\}\, a^k \bar{\omega}^h, \\ \frac{\partial \bar{\omega}^0_i}{\partial t} &= -\{(R^0_{ikh})_0 + (R^0_{ikh|l})_0\, a^l t + \cdots\}\, a^k \bar{\omega}^h, \end{aligned}\right. \tag{16}$$

where the a^i, δa^i are regarded as constants, has a solution $\bar{\omega}^i(t)$, $\bar{\omega}^j_i(t)$, $\bar{\omega}^0_i(t)$ satisfying the initial conditions

$$\bar{\omega}^i(0) = \bar{\omega}^j_i(0) = \bar{\omega}^0_i(0) = 0.$$

The components ω^i, ω^j_i, ω^0_i defined above are then obtained from $\bar{\omega}^i$, $\bar{\omega}^j_i$, $\bar{\omega}^0_i$ by putting $t = 1$ and by replacing the a^i by u^i. It follows that the condition (15) is equivalent to the condition

$$[\bar{\omega}^1 \cdots \bar{\omega}^n] = t^n[\delta a^1 \cdots \delta a^n]. \tag{17}$$

Put

$$\theta^i = a^k \bar{\omega}^i_k,$$
$$\bar{\omega}^0 = a^k \bar{\omega}^0_k.$$

[6] Cartan, Leçons, p. 223.

The system (16) gives

$$
(18) \quad \begin{cases}
\dfrac{\partial \bar{\omega}^i}{\partial t} = \delta a^i + \theta^i - \{(R^i_{0kh})_0\, a^k + (R^i_{0kh|l})_0\, a^k a^l t + \cdots\}\, \bar{\omega}^h, \\[2ex]
\dfrac{\partial \theta^i}{\partial t} = -2a^i \bar{\omega}^0 - \{(R^i_{jkh})_0\, a^j a^k + (R^i_{jkh|l})_0\, a^j a^k a^l t + \cdots\}\, \bar{\omega}^h, \\[2ex]
\dfrac{\partial \bar{\omega}^0}{\partial t} = -\{(R^0_{jkh})_0\, a^j a^k + (R^0_{jkh|l})_0\, a^j a^k a^l t + \cdots\}\, \bar{\omega}^h
\end{cases}
$$

In order that the normal coördinates of Cartan coincide with those of Veblen and Thomas at the point O it is necessary and sufficient that the system of equations (17), (18) has a solution $\bar{\omega}^i(t)$, $\theta^i(t)$, $\bar{\omega}^0(t)$ satisfying the initial conditions

$$\bar{\omega}^i(0) = \theta^i(0) = \bar{\omega}^0(0) = 0.$$

By expanding $\bar{\omega}^i$ according to powers of t

$$\bar{\omega}^i = \delta a^i t + \frac{1}{2!}\left(\frac{\partial^2 \bar{\omega}^i}{\partial t^2}\right)_0 t^2 + \cdots$$

from the equations (18) and substituting in (17), we get, on equating the coefficients of t^{n+1}, t^{n+2}, t^{n+3}, t^{n+4}, the relations

$$(R^i_{0ki})_0\, a^k = 0,$$

$$\{-4(R^i_{jki})_0 + (R^i_{0jp})_0\, (R^p_{0ki})_0\}\, a^j a^k = 0,$$

$$\{-2(R^i_{jki|l})_0 + (R^i_{0jp})_0\, (R^p_{0ki|l})_0\}\, a^j a^k a^l = 0,$$

$$
\begin{aligned}
\{4(R^i_{0jp|k})_0\, (R^p_{0li|m})_0 &+ 16(R^i_{jkp})_0\, (R^p_{0li|m})_0 - 4(R^i_{0jp|k})_0\, (R^p_{0lq})_0\, (R^q_{0mi})_0 \\
&- 8(R^i_{jkp})_0\, (R^p_{0lq})_0\, (R^q_{0mi})_0 + (R^i_{0jp})_0\, (R^p_{0kq})_0\, (R^q_{0lr})_0\, (R^r_{0mi})_0 \\
&+ 16(R^i_{jkp})_0\, (R^p_{lmi})_0\}\, a^j a^k a^l a^m = 0.
\end{aligned}
$$

These relations must hold for all values of a^i and since we consider the spaces for which the two kinds of normal coördinates coincide at every point, we can drop the last subscript 0. We get thus the necessary conditions

$$
(19) \quad \begin{cases}
R^i_{0ki} = 0, \\[1ex]
\displaystyle\sum_{(jk)} (-4R^i_{jki} + R^i_{0jp}\, R^p_{0ki}) = 0, \\[1ex]
\displaystyle\sum_{(jklm)} (4R^i_{0jp|k}\, R^p_{0li|m} + 16R^i_{jkp}\, R^p_{0li|m} - 4R^i_{0jp|k}\, R^p_{0lq}\, R^q_{0mi} \\
\qquad -8R^i_{jkp}\, R^p_{0lq}\, R^q_{0mi} + R^i_{0jp}\, R^p_{0kq}\, R^q_{0lr}\, R^r_{0mi} + 16R^i_{jkp}\, R^p_{lmi}) = 0,
\end{cases}
$$

where the summations are extended over all the permutations of j, k and j, k, l, m respectively.

When the space is without torsion, the expansion has been carried out to t^{n+6} inclusive and the following necessary conditions are found:

$$
(20)\quad \begin{cases} R^i_{jki} + R^i_{kji} = 0, \\ \sum_{(pqrs)} R^i_{pqj} R^j_{rsi} = 0, \\ \sum_{(lmpqrs)} (9R^i_{lmj|p} R^j_{qri|s} - 32 R^i_{lmj} R^j_{pqk} R^k_{rsi}) = 0. \end{cases}
$$

Unfortunately, there is no reason to expect that the conditions (19) or (20) be sufficient for general values of n. For $n \geqq 3$ it is clear that the condition that the space be with normal connection is not sufficient.

As Professor Cartan communicated to me, he has arrived at the necessary and sufficient conditions in the case $n = 2$. In this special case, in fact, the first two equations of (19) give

$$
(21)\qquad R^1_{012} = R^2_{012} = 0, \qquad R^2_{112} = R^1_{212} = 0, \qquad R^1_{112} = R^2_{212}.
$$

To show that the conditions (21) are also sufficient, it is only necessary to differentiate the equation (17) five times and take account of the equations (18). The fifth equation obtained is a consequence of the preceding ones. Therefore there exists a system of solutions satisfying (17), (18) and the given initial conditions. Consequently, in order that the normal coördinates of Cartan and those of Veblen and Thomas coincide at every point of a projectively connected space of two dimensions it is necessary and sufficient that the conditions (21) be satisfied. In this case, the condition that the connection be normal is sufficient. Geometrically, the conditions (21) signify that the infinitesimal displacement associated with any infinitely small cycle with origin A leaves invariant the point A and all the directions through A.

TSING-HUA UNIVERSITY,
PEIPING, CHINA.

ANNALS OF MATHEMATICS
Vol. 39, No. 1, January, 1938

RELATIONS BETWEEN RANKS OF A GENERAL MATRIX

BY RUFUS OLDENBURGER

(Received December 3, 1936)

1. **Introduction.** A *p-way matrix* $A = (a_{ij\cdots s})$ of *order* n is a matrix with p indices $i, j, \cdots, s$ and such that these indices have the range $1, \cdots, n$. With A are associated ranks[1] which are positive integers invariant under non-singular linear transformations

$$x_i = b_{i\alpha}x'_\alpha, \qquad y_j = c_{j\beta}y'_\beta, \cdots, z_s = d_{s\gamma}z'_\gamma,$$
$$i, \cdots, s, \qquad \alpha, \cdots, \gamma = 1, \cdots, n,$$

on the form

$$a_{ij\cdots s}x_iy_j \cdots z_s$$

associated with A. Some dependence relations between the ranks of a given p-way matrix, where $p \geqq 3$, were proved in a memoir[2] by the author. In particular it was shown that if a certain rank of a 3-way matrix A is equal to one, another rank of A is one. In the present paper the author proves that if the ranks of certain subsets S' of the set S of ranks of a p-way matrix A, $p \geqq 3$, are equal to one at least one of the ranks of S not contained in S' is equal to one.

Ranks which are equal to one are of particular interest because a p-way matrix A is the direct product of two or more matrices of lower dimension than p if and only if certain ranks of A equal one.

Applications of the theory to ordinary forms of any degree are given in a paragraph at the end of the paper.

It is assumed that the elements of A belong to any given field of numbers and that these elements are not all zero.

2. **Definitions.** An ordered set of indices $\tau = pq \cdots r$ of the indices of A is called[3] the *partition* τ. The partition τ is said to contain the indices $p, q, \cdots, r$; written $\tau \supset p, q, \cdots, r$. If the indices in τ are assigned numerical values $p', q', \cdots, r'$, then τ is said to have the *value* $\tau' = p'q' \cdots r'$.

Let $\alpha, \beta, \cdots, \psi$ be mutually exclusive, exhaustive partitions of the indices $i, j, \cdots, s$ of A. The element $a_{i'j'\cdots s'}$ will be denoted by $a_{\alpha'\beta'\cdots\psi'}$ where $\alpha, \beta, \cdots, \psi$ are understood to have the values $\alpha', \beta', \cdots, \psi'$ respectively,

---

[1] F. L. Hitchcock, *Multiple invariants and generalized rank of a p-way matrix or tensor*, Journal of Mathematics and Physics, vol. 7 (1927), pp. 40–79.

[2] R. Oldenburger, *Composition and rank of n-way matrices and multilinear forms*, Annals of Mathematics, vol. 35 (1934), pp. 641–643, 649–650.

[3] R. Oldenburger, Op. Cit., p. 623.

obtained by setting $i, j, \cdots, s$ respectively equal to $i', j', \cdots, s'$ in these partitions.

Let α', α'' denote two values of α; β', β'' two values of β; etc. Consider the relation

$$(1) \qquad \sum_{\gamma, \cdots, \psi} \begin{vmatrix} a_{\alpha'\beta'\gamma'\cdots\psi'} & a_{\alpha'\beta''\gamma''\cdots\psi''} \\ a_{\alpha''\beta'\gamma'\cdots\psi'} & a_{\alpha''\beta''\gamma''\cdots\psi''} \end{vmatrix} = 0,$$

where the summation is over all permutations of the pairs of values in the sets $(\gamma', \gamma''), \cdots, (\psi', \psi'')$. If (1) is satisfied for all choices of $\alpha', \alpha'', \cdots, \psi', \psi''$ not necessarily distinct the rank[4] $r[\alpha\beta, \gamma \cdots \psi]$ of A is said to be one. If a relation of type (1) is not satisfied the rank $r[\alpha\beta, \gamma \cdots \psi] \geqq 2$. In particular, if one of the partitions $\alpha, \beta \supset$ only one index, $r[\alpha\beta, \gamma \cdots \psi] \leqq n$, where n is the order of A. If in (1) the partitions $\gamma, \cdots, \psi$ do not occur, there is only one term in the left member of (1) and the associated rank is $r[\alpha\beta,]$. The value of $r[\alpha\beta, \gamma \cdots \psi]$ does not depend on the ordering of the indices in the respective partitions $\alpha, \cdots, \psi$. If we set $\gamma' = \gamma'', \cdots, \psi' = \psi''$ in (1) we obtain

LEMMA 1. *If the rank $r[\alpha\beta, \gamma \cdots \psi]$ of $A = (a_{ij\cdots s})$ is 1, the matrix*

$$\begin{pmatrix} a_{\alpha'\beta'\gamma'\cdots\psi'} & a_{\alpha'\beta''\gamma'\cdots\psi'} \\ a_{\alpha''\beta'\gamma'\cdots\psi'} & a_{\alpha''\beta''\gamma'\cdots\psi'} \end{pmatrix}$$

is singular for every choice of the values $\alpha', \alpha'', \beta', \beta'', \gamma', \cdots, \psi'$.

We also note the obvious

LEMMA 2. *If the matrix*

$$\begin{pmatrix} a_{\alpha'\beta'\gamma'\cdots\psi'} & a_{\alpha'\beta''\gamma''\cdots\psi''} \\ a_{\alpha''\beta'\gamma'\cdots\psi'} & a_{\alpha''\beta''\gamma''\cdots\psi''} \end{pmatrix}$$

is singular for every choice of $\alpha', \alpha'', \cdots, \psi', \psi''$, the rank $r[\alpha\beta, \gamma \cdots \psi]$ of A is 1.

3. **Theorems on ranks.** We shall prove

THEOREM 1. *Let a set S' of the ranks of $A = (a_{ijkm\cdots s})$ be defined by the property: for every index in the set $j \cdots s$ there exists a rank $r[\alpha\beta, \gamma \cdots \psi]$ in S' such that $\beta \supset$ this index and $\alpha \supset i$. If the ranks of S' are all equal to 1, the rank $r[i\phi,] = 1$, where $\phi = j \cdots s$.*

Let $\rho, \sigma, \tau, \cdots, \zeta$ be partitions such that $\rho \supset i$, $\sigma \supset j$ and $r[\rho\sigma, \tau \cdots \zeta] = 1$. Let $i', \cdots, s'$ be a set of values of $i, \cdots, s$. Let i'', j'' be another pair of values of i, j not necessarily distinct from i', j'. By Lemma 1 and $r[\rho\sigma, \tau \cdots \zeta] = 1$ we have

$$(2) \qquad \begin{pmatrix} a_{i'j'k'\cdots s'} & a_{i'j''k'\cdots s'} \\ a_{i''j'k'\cdots s'} & a_{i''j''k'\cdots s'} \end{pmatrix}$$

singular.

---

[4] R. Oldenburger, Op. Cit., p. 633.

Now let $\lambda, \mu, \nu, \cdots, \xi$ be partitions such that $\lambda \supset i, \mu \supset k$, and $r[\lambda\mu, \nu \cdots \xi] = 1$. Let k'' be a value of k not necessarily distinct from k'. By Lemma 1 and $r[\lambda\mu, \nu \cdots \xi] = 1$ we have

$$(3) \qquad \begin{pmatrix} a_{i'j''k'm'\cdots s'} & a_{i'j''k''m'\cdots s'} \\ a_{i''j''k'm'\cdots s'} & a_{i''j''k''m'\cdots s'} \end{pmatrix}$$

singular.

By (2) and (3) the matrix

$$\begin{pmatrix} a_{i'j'k'm'\cdots s'} & a_{i'j''k''m'\cdots s'} \\ a_{i''j'k'm'\cdots s'} & a_{i''j''k''m'\cdots s'} \end{pmatrix}$$

is singular. Thus by induction we obtain

$$(4) \qquad \begin{pmatrix} a_{i'j'\cdots s'} & a_{i'j''\cdots s''} \\ a_{i''j'\cdots s'} & a_{i''j''\cdots s''} \end{pmatrix}$$

singular for all possible values of the subscripts. By Lemma 2, $r[i\phi,] = 1$.

By a theorem[5] of another paper a rank $r = r[\alpha\beta, \gamma \cdots \psi]$, where $\alpha \equiv i$, satisfies the relation

$$r \leqq r[i\phi,].$$

Hence we have the

Corollary. *Let a set S' of ranks of $A = (a_{ij\cdots s})$ be defined as in Theorem 1 where $\alpha \equiv i$. The ranks of the set S' equal 1 if and only if $r[i\phi,] = 1$.*

We shall prove

Theorem 2. *Let a set S' of the ranks of $A = (a_{ij\cdots hl\cdots s})$ satisfy the properties:*

a.) For every index in the set $l, \cdots, s$ there exists a rank $r[\alpha\beta, \gamma \cdots \psi]$ in S' such that $\beta \supset$ this index and $\alpha \supset i$.

b.) For every pair of indices f in the set $j, \cdots, h$, and g in the set $l, \cdots, s$ there exists a rank $r[\delta\phi, \theta \cdots \zeta]$ in the set S' such that $\delta \supset f, \phi \supset g$.

c.) The set S' contains a rank $R = r[\rho\sigma, \tau \cdots \mu]$ such that $\rho = ij \cdots h$.

If all of the ranks in the set S' except R equal 1, then $R = 1$.

Denote the partitions ρ, σ' of $R' = r[\rho\sigma',]$ by

$$\rho = ijkf \cdots h, \qquad \sigma' = lmg \cdots s.$$

It follows by the same sort of argument as was used to obtain the singularity of (4), that the matrix

$$(4') \qquad \begin{pmatrix} a_{i'j'\cdots h'l'\cdots s'} & a_{i'j'\cdots h'l''\cdots s''} \\ a_{i''j'\cdots h'l'\cdots s'} & a_{i''j'\cdots h'l''\cdots s''} \end{pmatrix}$$

---

[5] R. Oldenburger, Op. Cit., p. 641.

is singular. If $\rho \equiv i$, the theorem is now proved. Assume henceforth that ρ contains more than one index.

Let $\alpha, \beta, \gamma, \cdots, \psi$ be partitions such that $\alpha \supset j$, $\beta \supset b$, where b is in the set $l, \cdots, s$, and $r[\alpha\beta, \gamma \cdots \psi] = 1$. This is possible by the property (b). By Lemma 1 this implies that the matrix

$$\begin{pmatrix} a_{i'j'k'f'\cdots l'\cdots a'b'c'\cdots s'} & a_{i'j''k'f'\cdots l'\cdots a'b'c'\cdots s'} \\ a_{i'j'k'f'\cdots l'\cdots a'b''c'\cdots s'} & a_{i'j''k'f'\cdots l'\cdots a'b''c'\cdots s'} \end{pmatrix} \tag{5}$$

is singular for all values $i = i'$; $j = j', j''$; $\cdots$; $s = s'$. Again if ρ contains 3 or more indices by property (b) there exists a rank $r[\lambda\mu, \nu \cdots \xi] = 1$ such that $\lambda \supset k$, $\mu \supset b$. By Lemma 1 the matrix

$$\begin{pmatrix} a_{i'j''k'f'\cdots l'\cdots a'b'c'\cdots s'} & a_{i'j''k''f'\cdots l'\cdots a'b'c'\cdots s'} \\ a_{i'j''k'f'\cdots l'\cdots a'b''c'\cdots s'} & a_{i'j''\kappa''f'\cdots l'\cdots a'b''c'\cdots s'} \end{pmatrix} \tag{6}$$

is singular. By (5) and (6) the matrix

$$\begin{pmatrix} a_{i'j'k'f'\cdots l'\cdots a'b'c'\cdots s'} & a_{i'j''k''f'\cdots l'\cdots a'b'c'\cdots s'} \\ a_{i'j'k'f'\cdots l'\cdots a'b''c'\cdots s'} & a_{i'j''k''f'\cdots l'\cdots a'b''c'\cdots s'} \end{pmatrix}$$

is singular. Thus by an induction process we obtain the singularity of

$$\begin{pmatrix} a_{i'j'\cdots h'l'\cdots a'b'c'\cdots s'} & a_{i'j''\cdots h''l'\cdots a'b'c'\cdots s'} \\ a_{i'j'\cdots h'l'\cdots a'b''c'\cdots s'} & a_{i'j''\cdots h''l'\cdots a'b''c'\cdots s'} \end{pmatrix}. \tag{7}$$

Transposing and taking $b \equiv l$, we get

$$\begin{pmatrix} a_{i'j'\cdots h'l'm'\cdots s'} & a_{i'j'\cdots h'l''m'\cdots s'} \\ a_{i'j''\cdots h''l'm'\cdots s'} & a_{i'j''\cdots h''l''m'\cdots s'} \end{pmatrix}. \tag{8}$$

If the set $l, \cdots, s$ contains at least two indices take $b \equiv m$ and $l = l''$ and obtain from (7) the singularity of

$$\begin{pmatrix} a_{i'j'\cdots h'l''m'g'\cdots s'} & a_{i'j'\cdots h'l''m''g'\cdots s'} \\ a_{i'j''\cdots h''l''m'g'\cdots s'} & a_{i'j''\cdots h''l''m''g'\cdots s'} \end{pmatrix}, \tag{9}$$

whence by (8) and (9)

$$\begin{pmatrix} a_{i'j'\cdots h'l'm'g'\cdots s'} & a_{i'j'\cdots h'l''m''g'\cdots s'} \\ a_{i'j''\cdots h''l'm'g'\cdots s'} & a_{i'j''\cdots h''l''m''g'\cdots s'} \end{pmatrix}$$

is singular. Thus by induction we obtain the singularity of

$$\begin{pmatrix} a_{i'j'\cdots h'l'\cdots s'} & a_{i'j'\cdots h'l''\cdots s''} \\ a_{i'j''\cdots h''l'\cdots s'} & a_{i'j''\cdots h''l''\cdots s''} \end{pmatrix} \tag{10}$$

for all values of the subscripts.

Setting $j' = j''$, $\cdots$, $h' = h''$ in (4') we obtain

$$\begin{pmatrix} a_{i'j''\cdots h''l'\cdots s'} & a_{i'j''\cdots h''l''\cdots s''} \\ a_{i''j''\cdots h''l'\cdots s'} & a_{i''j''\cdots h''l''\cdots s''} \end{pmatrix}. \tag{11}$$

By (10) and (11) we obtain the singularity of

$$\begin{pmatrix} a_{i'j'\cdots h'l'\cdots s'} & a_{i'j'\cdots h'l''\cdots s''} \\ a_{i''j''\cdots h''l'\cdots s'} & a_{i''j''\cdots h''l''\cdots s''} \end{pmatrix}$$

for all values of the subscripts. By Lemma 2, $R' = 1$. By $R \leqq R'$, $R = 1$.

4. **Remarks on the weakening of the assumptions of Theorems 1 and 2.** That S' in Theorem 1 cannot in general be replaced by a subset S'' of S', where $S'' \not\equiv S'$ is apparent from the following example.

Consider $A = (a_{ij\cdots hl\cdots s})$ of order 2 where $a_{ij\cdots hl\cdots s} = 1$ for $(i, j, \cdots, h, l, \cdots, s) = (1, 1, \cdots, 1, 1, \cdots, 1), (2, 1, \cdots, 1, 2, \cdots 2)$, and all other elements of A vanish. Then

$$r[i\beta, \delta \cdots \psi] = 1$$

for $\beta = j, \cdots, h$ respectively, and

$$r[i\beta, \delta \cdots \psi] = r[i\phi,] = 2$$

for $\beta = l, \cdots, s$ respectively, and $\Phi = j \cdots s$.

That property (b) of Theorem 2 can not be omitted without changing the results of Theorem 2 is evident from the example where

$$a_{1111} = a_{1221} = 1, \tag{12}$$

and all other elements of a matrix $A = (a_{ijkl})$ of order 2 vanish. Then

$$r[(il)k, j] = r[(ik)l, j] = 1; r[(ij)k, l] = 2, \tag{13}$$

where (il), (ik), (ij) denote the partitions il, ik, ij respectively.

That property (a) cannot be omitted from Theorem 2 without changing the results is evident from the example of a matrix of order 2 whose only non-vanishing elements are

$$a_{1111} = a_{2121} = 1. \tag{14}$$

Then

$$r[(ik)j, l] = r[(il)j, k] = 1; \qquad r[(ij)k, l] = 2. \tag{15}$$

It can be shown that the above examples also prove that conditions (a) or (b) of Theorem 2 cannot be weakened in general to proper subsets of the ranks defined by (a) or (b).

5. **Application of Theorems 1 and 2 to forms and symmetric matrices.** A *symmetric* matrix $a = (a_{ij\cdots m})$ is a matrix whose elements satisfy the property

$$a_{ij\cdots m} = a_{i_1 j_1 \cdots m_1},$$

where $i_1, j_1, \cdots, m_1$ are obtained from $i, j, \cdots, m$ by any substitution on $i, j, \cdots, m$. Thus

$$a_{1\cdots 12} = a_{1\cdots 121} = \cdots = a_{21\cdots 1}; \qquad a_{1\cdots 122} = a_{1\cdots 212} = \cdots = a_{221\cdots 1}; \text{ etc.}$$

With any n-ary form F of degree p, whose elements belong to a field ϕ not having characteristic $2, 3, \cdots, p!$, can be associated a unique symmetric p-way matrix $(a_{ij\cdots m})$, where $i, j, \cdots, m$ range over $1, \cdots, n$, and

$$F = a_{ij\cdots m} x_i x_j \cdots x_m .$$

By symmetry a rank $r(\alpha\beta, \gamma \cdots \psi)$ of A depends only on the number of indices in the partitions $\alpha, \cdots, \psi$, and not upon the choice of indices in these partitions. By a theorem of another paper[6] for α sufficiently restricted

$$r[\alpha\beta, \gamma \cdots \psi] \leqq r[i\phi,]. \tag{16}$$

By Theorem 1 if any rank $r(\alpha\beta, \gamma \cdots \psi) = 1$, $r(i\phi,) = 1$. In this case (16) becomes an equality. We have proved

THEOREM 3. *If any rank $r(\alpha\beta, \gamma \cdots \psi)$ of a symmetric matrix $A = (a_{ij\cdots m})$ is* 1, *the matrix A is a direct product of vectors and the associated form F can be written as a product of linear forms, $a_i x_i, b_j x_j, \cdots, d_m x_m$.*

In another paper the author defined "signant" ranks[7] of a matrix. These include the ranks of this paper. If the conditions of Theorem 3 are satisfied, it can be shown that any signant rank of A is 1.

If $r[\alpha\beta, \gamma \cdots \psi] = 1$, parts a and b of Theorem 2 are satisfied for any choice of indices $j \cdots h$. Hence Theorem 2 implies that any rank $r[\rho\sigma, \tau \cdots \mu]$ of A is 1. This result was already obtained from Theorem 1.

We shall say that the 2 partitions α, β are signant in $r[\alpha\beta, \gamma \cdots \psi]$. We have proved

THEOREM 4. *For a binary form*

$$F = a_{ij\cdots m} x_i x_j \cdots x_m$$

of any degree the ranks with 2 *partitions signant are equal.*

ARMOUR INSTITUTE OF TECHNOLOGY.

[6] Oldenburger, Op. Cit., p. 641.
[7] See Oldenburger, Op. Cit., pp. 633–635.

Annals of Mathematics
Vol. 39, No. 1, January, 1938

THE CANONICAL FORM OF A MATRIX

By J. H. M. Wedderburn

(Received November 10, 1937)

If $\varphi(\lambda) = \lambda^n + a_1\lambda^{n-1} + \cdots + a_n$ is irreducible in a field F, it is well known that we may take as the canonical form of a matrix of order n for which φ is the reduced characteristic function

$$(1) \qquad A = \begin{Vmatrix} -a_1 & -a_2 & \cdots & -a_{n-1} & -a_n \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdots & \cdot & \cdot \\ 0 & 0 & \cdots & 1 & 0 \end{Vmatrix}$$

and also, subject to certain restrictions on the field or on φ, a matrix of order nr whose characteristic function is $[\varphi(\lambda)]^r$ is given by

$$(2) \qquad B = \begin{Vmatrix} A & 1_n & & & \\ & A & 1_n & & \\ & & \ddots & \ddots & \\ & & & \cdot & 1_n \\ & & & & A \end{Vmatrix} \qquad (r \text{ rows})$$

where 1_n is the identity matrix of order n and coördinates not indicated are 0. If the terms in (2) are rearranged by taking the (i, j) coördinate of each block of terms to form a new block, we get the equivalent matrix

$$(3) \qquad C = \begin{Vmatrix} -a_{1r} + k_r & -a_{2r} & \cdots & -a_{n-1,r} & -a_{nr} \\ 1_r & k_r & \cdots & 0 & 0 \\ \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdots & \cdot & \cdot \\ 0 & 0 & \cdots & k_r & 0 \\ 0 & 0 & \cdots & 1_r & k_r \end{Vmatrix} \qquad (n \text{ rows})$$

where k_r is the nilpotent matrix of order and index r in which the coördinates of the diagonal immediately to the right of the main diagonal equal 1 and all other coördinates are 0; and a_{ir} is the scalar matrix of order r corresponding to a_i.

We may also write (3) in the form

$$C = C_1 + K$$

where C_1 is formed in the same way as A in (1) except that the coördinates are scalar matrices of order r; and K has k_r in the main diagonal and 0 elsewhere so that it is a nilpotent matrix of order nr and index r. Since the matrix coördinates of C are commutative, we have the integral identity

$$\varphi(C) = \varphi(C_1) + K\varphi'(C_1) + \cdots.$$

But $\varphi(C_1) = 0$ and $\varphi'(C_1) \neq 0$ unless $\varphi'(\lambda) \equiv 0$; hence, if $\varphi'(\lambda) \not\equiv 0$, the r^{th} power of $\varphi(C)$ is the first which is 0, that is, the reduced characteristic function of C is $[\varphi(\lambda)]^r$ and (2) and (3) give available rational canonical forms for C.

The exceptional case, $\varphi'(\lambda) \equiv 0$, can only occur if the characteristic of F is $p > 0$ and $\varphi(\lambda)$ is a polynomial in λ^p. The following short discussion gives a canonical form which remains valid even in this exceptional case; it is a rearrangement of one communicated orally to me by my friend Dr. N. Jacobson.

Let D be the matrix of order nr

$$D = \begin{Vmatrix} -d_1 & -d_2 & \cdots & -d_{n-1} & -d_n \\ 1_r & 0 & \cdots & 0 & 0 \\ \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdots & \cdot & \cdot \\ 0 & 0 & \cdots & 1_r & 0 \end{Vmatrix} \qquad (n \text{ rows}) \tag{4}$$

where d_i is any matrix of order r; also let D_i be the matrix with d_i in the main diagonal and 0 elsewhere. Then the same method as is used in showing that $\varphi(A) = 0$ in (1) shows that

$$D^n + D_1 D^{n-1} + \cdots + D_{n-1}D + D_n = 0 \tag{5}$$

when the d_i are commutative and, when this is not assumed, it is easily seen that (5) follows from the recursion formula for the coefficients of D^{r+1} in terms of those of D^r. Suppose now that $d_i = a_{ir}$ $(i = 1, 2, \cdots, n-1)$, $d_n = a_{nr} - h_r$ where the notation for a_{ir} is the same as in (3) and h_r is at first arbitrary and is later taken to be the same as k_r, that is, a nilpotent matrix of order and index r; further let A_i and H be formed from a_{ir} and h_r in the same way as D_i from d_i. We have then on using (5)

$$\begin{aligned} \varphi(D) &= D^n + A_1 D^{n-1} + \cdots + A_{n-1}D + A_n \\ &= D^n + \cdots + A_{n-1}D + A_n - H + H \\ &= H \end{aligned}$$

and, if we take H to be nilpotent with the index r, it follows that $[\varphi(D)]^r = 0$, $[\varphi(D)]^{r-1} \neq 0$. Hence the reduced characteristic function of D is $[\varphi(\lambda)]^r$.

If we change D in the same way as in passing from C to B, we get

$$(6)\qquad \begin{Vmatrix} A & g_n & & & & \\ & A & g_n & & & \\ & & A & \cdot & & \\ & & & \cdot & \cdot & \\ & & & & A & g_n \\ & & & & & A \end{Vmatrix} \qquad (r \text{ rows})$$

where g_n is the matrix of order n all of whose coördinates are 0 except the one in the position $(1, n)$ which is 1. This is Jacobson's canonical form. By modifying a different a_{ir} we may clearly take the coördinate which is not 0 to be any one in the first row except $(1, 1)$.

Another, and perhaps better, form can be found as follows. Let the characteristic of F be $p \neq 0$ and suppose that $\varphi(\lambda) = \psi(\lambda^{p^m})$ is a polynomial in λ^{p^m} but not in $\lambda^{p^{m+1}}$; let

$$\psi(\lambda) = \lambda^s + a_1\lambda^{s-1} + \cdots + a^s$$

$$u_s = \begin{Vmatrix} -a_1 & -a_2 & \cdots & -a_{s-1} & -a_s \\ 1 & 0 & \cdots & 0 & 0 \\ \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdots & \cdot & \cdot \\ 0 & 0 & \cdots & 1 & 0 \end{Vmatrix} \qquad (s \text{ rows})$$

$$U = \begin{Vmatrix} u_s & 1_s & & & \\ & u_s & 1_s & & \\ & & \cdot & \cdot & \\ & & & u_s & 1_s \\ & & & & u_s \end{Vmatrix} \qquad (r \text{ rows})$$

$$(7)\qquad V = \begin{Vmatrix} 0 & 0 & \cdots & 0 & U \\ 1_{rs} & 0 & \cdots & 0 & 0 \\ 0 & 1_{rs} & \cdots & 0 & 0 \\ \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdots & \cdot & \cdot \\ 0 & 0 & \cdots & 1_{rs} & 0 \end{Vmatrix} \qquad (p^m \text{ rows}).$$

It then follows as above that the characteristic function of V is $[\varphi(\lambda)]^r$. When the characteristic of F is 0, the only change is that V reduces to U and ψ to φ. Form (7) is more convenient than (6) when functions of the canonical matrix are under consideration.

PRINCETON UNIVERSITY.

Annals of Mathematics
Vol. 39, No. 1, January, 1938

SIMPLE LIE ALGEBRAS OF TYPE A[1]

By N. Jacobson[2]

(Received March 24, 1937)

In a recent paper[3] the author discussed the Lie algebras of characteristic 0 obtainable as the set of skew elements of an involutorial normal simple associative algebra. The present paper gives an extension of these results to simple associative algebras of second kind as defined by Albert.[4] The resulting Lie algebras in this case constitute Landherr's class A_{II}.[5] We reduce the problem of classifying these Lie algebras, as in our earlier work on Lie algebras of types B, C, D,[6] to two standard problems in associative algebra, namely, classification of involutorial simple algebras and of generalized hermitian matrices relative to cogredience. In this sense a complete determination of normal simple Lie algebras of characteristic 0 except those of a finite number of orders[7] results from Landherr's work on the algebras of type A_{I} and our own on types A_{II}, B, C, D.

1. Let $\mathfrak{A}$ be a simple associative algebra of order $2n^2$ over the field Φ of characteristic 0 and $\mathrm{P} = \Phi(q)$, $q^2 = \mu \in \Phi$, be its centrum. Suppose $\mathfrak{A}$ is involutorial of the second kind i.e. there is defined a $(1 - 1)$ correspondence J in $\mathfrak{A}$ such that

$$(a + b)^J = a^J + b^J \qquad (ab)^J = b^J a^J \qquad (\xi a)^J = \xi^J a^J \equiv \bar{\xi} a^J \qquad a^{J^2} = a$$

where $\xi^J \equiv \bar{\xi} = \alpha - \beta q$ if $\xi = \alpha + \beta q$, $\alpha, \beta \in \Phi$. We shall call J an involutorial anti-automorphism (i.a.a.) of *second kind*. It is easily seen that $\mathfrak{S}_J$ the set of J-skew elements $(a^J = -a)$ forms a Lie algebra over Φ relative to commutator multiplication $[a, b] \equiv ab - ba$.[8] $\mathfrak{H}_J$ the set of J-symmetric elements $(a^J = a)$ is a vector space over Φ and if $a \in \mathfrak{H}_J$, $qa \in \mathfrak{S}_J$. On the other

[1] Presented to the Society March 26, 1937. An abstract of this paper appeared in the Proc. Nat. Acad. Sci. **23** (1937) pp. 240–242.

[2] National Research Fellow.

[3] *A class of normal simple Lie algebras of characteristic* 0, Annals of Math. **38** (1937) pp. 508-517, referred to hereafter as *J*.

[4] A. A. Albert, *Involutorial simple algebras and real Riemann matrices*, Annals of Math. **36** (1935) pp. 894-910.

[5] W. Landherr, *Über einfache Liesche Ringe*, Hamb. Abhandlungen **11** (1935) pp. 41-64, referred to as *L*.

[6] J.

[7] The exceptional orders are those of the five exceptional Lie algebras of Killing and Cartan and order 28 corresponding to an algebra of type *D*. Cf. E. Cartan, *Thèse* Paris 1894, Chap. 5 and J. p. 513. The Lie algebras of type A_{II} over a p-adic field were determined prior to the present work by Landherr loc. cit.

[8] Cf. J. p. 508.

hand if $a \in \mathfrak{S}_J$, $qa \in \mathfrak{H}_J$. Thus the two vector spaces $\mathfrak{S}_J$ and $\mathfrak{H}_J$ have the same order over Φ. Evidently the intersection $\mathfrak{S}_J \frown \mathfrak{H}_J = 0$ and if a is any element of $\mathfrak{A}$, $a = \frac{1}{2}(a + a^J) + \frac{1}{2}(a - a^J) = b + c$ where $b \in \mathfrak{H}_J$ and $c \in \mathfrak{S}_J$. Thus $\mathfrak{A} = \mathfrak{S}_J + \mathfrak{H}_J$ and hence the order of $\mathfrak{S}_J$ over Φ, $(\mathfrak{S}_J : \Phi) = (\mathfrak{H}_J : \Phi) = n^2$. If $a_0, a_1, \cdots, a_m$ $(m = n^2 - 1)$ is a basis for $\mathfrak{S}_J$ over Φ, $qa_0, qa_1, \cdots, qa_m$ is one for $\mathfrak{H}_J$ over Φ and so $a_0, a_1, \cdots, a_m$ is a basis for $\mathfrak{A}$ over P. Hence $\mathfrak{S}_{JP}$[9] is isomorphic to $\mathfrak{A}$ over P regarded as a Lie algebra relative to commutator multiplication. If $\mathfrak{B}$ is any Lie algebra we denote its derived algebra $[\mathfrak{B}, \mathfrak{B}]$ by $\mathfrak{B}'$. Evidently $(\mathfrak{B}_\Lambda)' = (\mathfrak{B}')_\Lambda$ for Λ any extension field of the field of $\mathfrak{B}$. In particular $\mathfrak{S}'_{JP} \cong \mathfrak{A}'$ (over P).

If Ω is the algebraic closure of P (or of Φ), for $\mathfrak{A}$ over P we have $\mathfrak{A}_\Omega \cong \Omega_n$ the matrix algebra of n rows and columns over Ω and hence $\mathfrak{S}'_{J\Omega} \cong \Omega'_n$.[10] Since Ω'_n is simple,[11] $\mathfrak{S}'_J$ is a normal simple Lie algebra. We note also that $\mathfrak{S}'_J$ has order m over Φ since Ω'_n has this order over Ω and we may suppose that in the basis $a_0, a_1, \cdots, a_m$ for $\mathfrak{A}$ over P, $a_1, \cdots, a_m$ form a basis for $\mathfrak{S}'_J$ over Φ. Since $[q, a] = 0$, $q \notin \mathfrak{S}'_J$ for otherwise $\mathfrak{S}'_J$ would not be simple. Hence we may suppose also that $a_0 = q$ in the basis $a_0, a_1, \cdots, a_m$ for $\mathfrak{A}$ over P or $\mathfrak{S}_J$ over Φ.

Considering $\mathfrak{A}$ over P, the condition $\mathfrak{A}_\Omega \cong \Omega_n$ means that $\mathfrak{A}$ over P may be represented by matrices in Ω_n such that the linear combinations of these matrices with coefficients in Ω is Ω_n. Hence if $a_i \to A_i$ in the representation $A_0, A_1, \cdots, A_m$ form a basis for Ω_n and $A_1, \cdots, A_m$ for Ω'_n over Ω. Employing a matric basis E_{ij} such that $E_{ij}E_{kl} = \delta_{jk}E_{il}$ we see that Ω'_n is generated by E_{ij} $(i \neq j)$ and $E_{ii} - E_{jj}$ and hence consists of all the matrices of trace 0 and so $A_1, \cdots, A_m$ is a basis over Ω for these matrices.

The enveloping algebra over Φ of $\mathfrak{S}_J$ (smallest algebra over Φ contained in $\mathfrak{A}$ and containing $\mathfrak{S}_J$) is $\mathfrak{A}$ itself since it contains $q, a_1, \cdots, a_m$ and $q^2, qa_1, \cdots, qa_m$. We shall require the stronger result:

LEMMA 1. *If $n > 2$ the enveloping algebra over Φ of $\mathfrak{S}'_J$ is $\mathfrak{A}$.*

Let $\mathfrak{B}$ denote this algebra. Since $\mathfrak{B}$ contains $a_1, \cdots, a_m$ it suffices to show that $\mathfrak{B}$ contains q also and by the above discussion $\mathfrak{B}$ will contain q if it contains any element of $\mathfrak{S}_J$ not in $\mathfrak{S}'_J$. Now there exists a matrix W in Ω_n such that $\mathrm{tr}\, W = 0$ but $\mathrm{tr}\, W^3 \neq 0$. For example we may take

$$W = \rho_1 E_{11} + \rho_2 E_{22} + \cdots + \rho_n E_{nn}$$

where $\rho_1 + \rho_2 + \cdots + \rho_n = 0$ but $\rho_1^3 + \rho_2^3 + \cdots + \rho_n^3 \neq 0$. Then $W = \omega_1 A_1 + \cdots + \omega_m A_m$, $\omega_i \in \Omega$. It follows that there are elements $\alpha_1, \cdots, \alpha_m$ in Φ also such that $\mathrm{tr}\, A^3 \neq 0$ where $A = \alpha_1 A_1 + \cdots + \alpha_m A_m$. Then the

---

[9] If $\mathfrak{A}$ is a Lie algebra or an associative algebra with basis $a_1, a_2, \cdots$ over Φ and Σ an extension field of Φ then $\mathfrak{A}_\Sigma$ denotes the algebra having the basis $a_1, a_2, \cdots$, but with coefficients in Σ.

[10] In general we shall denote the algebra of n-rowed matrices with coördinates in an algebra $\mathfrak{A}$ by $\mathfrak{A}_n$.

[11] Ω'_n is one of the simple algebras (class A) in Killing-Cartan's classification. If Ω is the field of complex numbers Ω'_n is the infinitesimal group of the unimodular linear group.

element $a = \alpha_1 a_1 + \cdots + \alpha_m a_m \in \mathfrak{S}_J'$ and a^3 is an element of $\mathfrak{B}$ contained in $\mathfrak{Z}_J$ but not in $\mathfrak{S}_J'$.

It is well known that the field of the coefficients of the characteristic equations of the matrices (representing the elements) of $\mathfrak{A}$ over P in Ω_n is P.[12] By Newton's identities this field is the same as the field of the traces of the matrices of $\mathfrak{A}$. It follows that if $n > 2$ the field Λ of the characteristic equations of the matrices of $\mathfrak{S}_J'$ is also P. For by a result of an earlier paper[13] the traces of the matrices of an enveloping algebra of a Lie algebra $\mathfrak{L}$ are expressible rationally in terms of the coefficients of the characteristic equations of the elements of $\mathfrak{L}$. Hence by Lemma 1 $\Lambda = \text{P}$. Following Landherr[14] we shall say that a Lie algebra $\mathfrak{L}$ over Φ has *type* A_{II} if $\mathfrak{L}_\Omega \cong \Omega_n'$ and in the representation of $\mathfrak{L}$ by matrices in Ω_n' the field of the characteristic equations of the matrices of $\mathfrak{L}$ is a proper overfield of Φ. With this definition we have

THEOREM 1. *If $\mathfrak{A}$ is an involutorial simple algebra of second kind and order $2n^2$, $n > 2$, over Φ and J is an* i.a.a. *of second kind in $\mathfrak{A}$ then $\mathfrak{S}_J'$ is a normal simple Lie algebra of type A_{II}.*

If $n = 2$ we have for any a in $\mathfrak{A}$ that $a^2 - \text{tr}\,(a)a + N(a) = 0$ where $N(a)$ is the determinant in the representation in Ω_n. Hence if $a \in \mathfrak{S}_J'$, $a^2 = \alpha$ and since a^2 is J-symmetric $\alpha \in \Phi$. Thus the field $\Lambda = \Phi$ in this case and the enveloping algebra over Φ of $\mathfrak{S}_J'$ is $\mathfrak{B} \neq \mathfrak{A}$.

2. If G is an automorphism of $\mathfrak{A}$ over Φ it induces an automorphism in P. Hence we have either $\xi^G = \xi$ for all ξ in P or $\xi^G = \bar{\xi}$. In the former case G is an automorphism of $\mathfrak{A}$ over P and hence is inner i.e. there exists an element g in $\mathfrak{A}$ such that $a^G = g^{-1}ag$ for all a.[15] If S is a second automorphism then $S^{-1}GS$ is the inner automorphism $a \to (g^S)^{-1}ag^S$. Thus the inner automorphisms constitute an invariant subgroup of the complete group of automorphisms. If S_1 and S_2 are outer automorphisms we must have $\xi^{S_1} = \bar{\xi} = \xi^{S_2}$ and hence $\xi^{S_1S_2^{-1}} = \xi$. It follows that $S_1S_2^{-1}$ is inner and so if there exist outer automorphisms the index of the group of inner automorphisms in the complete group is 2.

We note that if S is an outer automorphism and J an i.a.a. of second kind then $\xi^{SJ} = \xi$ and hence $V = SJ$ is an anti-automorphism of $\mathfrak{A}$ over P. Since $\mathfrak{A}$ over P is normal simple it follows from a theorem of Brauer's that $\mathfrak{A}$ over P has exponent 1 or 2 and if $\mathfrak{A} = \mathfrak{F}_r$ where $\mathfrak{F}$ is a normal division algebra then the degree of $\mathfrak{F}$ is 2^e, $e \geqq 0$.[16] Conversely if $\mathfrak{A}$ over P has exponent 1 or 2 it has an anti-automorphism V and hence $\mathfrak{A}$ over Φ has an outer automorphism $S = VJ$.

---

[12] See for example M. Deuring, *Algebren*, Ergebnisse der Math. Springer 1935, pp. 50–52.

[13] N. Jacobson, *Rational methods in the theory of Lie algebras*, Annals of Math. **36** (1935) pp. 879–880.

[14] L. p. 50.

[15] For a proof of this theorem see Deuring's *Algebren*, p. 43.

[16] R. Brauer, *Über Systeme hyperkomplexer Zahlen*, Math. Zeitsch. **30** (1929) p. 103 or Deuring's Algebren p. 45 and pp. 58–59.

If K is a second i.a.a. (not necessarily distinct from J) cogredient to J i.e. $K = S^{-1}JS$ for some automorphism S then K is also of the second kind and it is easily seen that $\mathfrak{S}_J \cong \mathfrak{S}_K = \mathfrak{S}_J^S$ and hence also $\mathfrak{S}'_J \cong \mathfrak{S}'_K$.[17] The following criterion for cogredience will be used below:

LEMMA 2. *Suppose $n > 2$ and J and K are i.a.a. of the second kind such that there is an automorphism S of $\mathfrak{A}$ over Φ carrying $\mathfrak{S}'_J$ into $\mathfrak{S}'_K$. Then $K = S^{-1}JS$.*

Since $\mathfrak{S}'_{S^{-1}JS} = (\mathfrak{S}'_J)^S = \mathfrak{S}'_K = \mathfrak{S}'$ we have for $a \in \mathfrak{S}'$ that $a^{S^{-1}JS} = -a = a^K$. But by Lemma 1 any element b of $\mathfrak{A}$ has the form $\sum \beta_{i_1 \cdots i_r} a_{i_1} a_{i_2} \cdots a_{i_r}$ $(i_\alpha = 1, \cdots, m;\ \beta \in \Phi)$ and hence

$$b^{S^{-1}JS} = \sum (-1)^r \beta_{i_1 \cdots i_r} a_{i_r} a_{i_{r-1}} \cdots a_{i_1} = b^K$$

and $K = S^{-1}JS$.

As a special case of Lemma 2 we note that if S is an automorphism mapping $\mathfrak{S}'_J$ into itself then $S^{-1}JS = J$ i.e. S commutes with J. The converse of this is immediate. If $S = G$ is inner, say $a^G = g^{-1}ag$, then $a^{G^{-1}JG} = (g^J g)^{-1} a^J (g^J g)$ and the condition $GJ = JG$ is that $g^J g = \gamma \in \mathrm{P}$. Then $gg^J = \gamma$ also and $(gg^J)^J = \bar{\gamma} = g^J g = \gamma \in \Phi$ i.e. g is a J-orthogonal element.[18]

3. The following lemma, proved by Landherr,[19] is fundamental for the remainder of the paper:

LEMMA 3. *Let $A \to A^G$ be an automorphism of the Lie algebra Φ'_n over Φ, then there exists a matrix G in Φ_n such that either $A^G = G^{-1}AG$ or $A^G = -G^{-1}A'G$ where A' is the transposed of A.*

The automorphisms which may be expressed in the form $A \to G^{-1}AG$ will be called *inner*. As in the preceding section we may show that they form an invariant subgroup $\mathfrak{G}_0$ of the complete group of automorphisms $\mathfrak{G}$. If $n = 2$ we have $-A' = Q^{-1}AQ$ where $Q = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ and A is any matrix of trace 0 i.e. in Φ'_n. Hence $\mathfrak{G}_0 = \mathfrak{G}$. On the other hand if $n > 2$ there exists no such Q since this condition implies tr $A^3 = 0$. Hence in this case $\mathfrak{G}_0 \neq \mathfrak{G}$ and it is easily seen that $\mathfrak{G}_0$ has index 2 in $\mathfrak{G}$.

Now suppose $\mathfrak{A}_1$ and $\mathfrak{A}_2$ are two involutorial simple algebras of order $2n^2$, $n > 2$ over Φ, $\mathrm{P}_1 = \Phi(q_1)$, $\mathrm{P}_2 = \Phi(q_2)$ their centrums and J_1 and J_2 i.a.a. of second kind such that the Lie algebras over Φ, $\mathfrak{S}'_{J_1}$ and $\mathfrak{S}'_{J_2}$ are isomorphic. By passing to isomorphic fields if necessary we may suppose that P_1 and P_2 have the same algebraic closure Ω. Then $\mathfrak{A}_1$ and $\mathfrak{A}_2$ have representations by matrices in Ω_n such that the linear combinations in Ω of the matrices of $\mathfrak{S}'_{J_1}$ or of $\mathfrak{S}'_{J_2}$ constitute Ω'_n. Let $a_0^{(1)}, a_1^{(1)}, \cdots, a_m^{(1)}$ be a basis of $\mathfrak{S}_{J_1}$ over Φ and hence of $\mathfrak{A}_1$ over P_1 such that $a_0^{(1)} = q_1$ and $a_1^{(1)}, \cdots, a_m^{(1)}$ is a basis of $\mathfrak{S}'_{J_1}$ over Φ.

---

[17] J. p. 509.

[18] J. p. 509.

[19] L. p. 59. Cf. A. Weinstein, *Fundementalsatz der Tensorrechnung*, Math. Zeitsch. **16** (1923) pp. 78-91.

If $a_1^{(2)}, \cdots, a_m^{(2)}$ are the elements corresponding to $a_1^{(1)}, \cdots, a_m^{(1)}$ in the isomorphism between $\mathfrak{S}'_{J_1}$ and $\mathfrak{S}'_{J_2}$ and $a_0^{(2)} = q_2$, then $a_i^{(2)}$, $i = 0, \cdots, m$ is a basis of $\mathfrak{S}_{J_2}$ over Φ and of $\mathfrak{A}_2$ over P_2. If $a_i^{(1)} \to A_i^{(1)}$ and $a_i^{(2)} \to A_i^{(2)}$ in the representations of $\mathfrak{A}_1$ and $\mathfrak{A}_2$ then the correspondence $\sum_{j=1}^m \omega_j A_j^{(1)} \to \sum \omega_j A_j^{(2)}$ where the ω's are arbitrary in Ω is an automorphism of Ω'_n over Ω. Hence by Lemma 3 there exists a matrix G in Ω_n such that either $A_j^{(2)} = G^{-1}A_j^{(1)}G$ or $A_j^{(2)} = -G^{-1}(A_j^{(1)})'G$. By Lemma 1 the matrices of $\mathfrak{A}_1$ and $\mathfrak{A}_2$ have the form $B_1 = \sum \beta_{i_1 \cdots i_r} A_{i_1}^{(1)} \cdots A_{i_r}^{(1)}$ and $B_2 = \sum \beta_{i_1 \cdots i_r} A_{i_1}^{(2)} \cdots A_{i_r}^{(2)}$ ($i_\alpha = 1, \cdots, m$; $\beta \in \Phi$) respectively. Thus if $A_j^{(2)} = G^{-1}A_j^{(1)}G$ then the correspondence G defined by $B_1^G = G^{-1}B_1G$ is an isomorphism between the matrices of $\mathfrak{A}_1$ and those of $\mathfrak{A}_2$ which extends the isomorphism $\sum \alpha_j A_j^{(1)} \to \sum \alpha_j A_j^{(2)}$ ($\alpha \in \Phi$) between the matrices of $\mathfrak{S}'_{J_1}$ and those of $\mathfrak{S}'_{J_2}$. On the other hand if $A_j^{(2)} = -G^{-1}(A_j^{(1)})'G$ then G defined by $B_1^G = (G^{-1}B_1'G)^{J_2}$ has this property. In any case $\mathfrak{A}_1$ and $\mathfrak{A}_2$ are isomorphic.

If $\mathfrak{A}_1$ and $\mathfrak{A}_2$ are identical, say $= \mathfrak{A}$, the argument shows that the isomorphism between $\mathfrak{S}'_{J_1}$ and $\mathfrak{S}'_{J_2}$ may be extended to an automorphism G of $\mathfrak{A}$ over Φ. Evidently the correspondence $B_1 \to G^{-1}B_1G$ and $B_1 \to G^{-1}B_1'G$ leave the matrices of P (scalar matrices) unaltered. Hence in the first case G is an inner automorphism of $\mathfrak{A}$ and in the second case it is outer. As noted earlier the latter possibility is excluded if the exponent of $\mathfrak{A}$ over P > 2. By Lemma 2 we have

THEOREM 2. *If $\mathfrak{A}_1$ and $\mathfrak{A}_2$ are involutorial simple algebras of order $2n^2$, $n > 2$, over Φ and J_1 and J_2 i.a.a. of second kind in $\mathfrak{A}_1$ and $\mathfrak{A}_2$ respectively such that $\mathfrak{S}'_{J_1} \cong \mathfrak{S}'_{J_2}$ then $\mathfrak{A}_1 \cong \mathfrak{A}_2$ and when $\mathfrak{A}_1$ and $\mathfrak{A}_2$ are identified J_1 and J_2 are cogredient.*

THEOREM 3. *If $\mathfrak{A}$ is involutorial simple of order $2n^2$, $n > 2$, over Φ and J an i.a.a. of second kind in $\mathfrak{A}$ then any automorphism in $\mathfrak{S}'_J$ may be extended to an automorphism of the associative algebra $\mathfrak{A}$.*

This results by noting that the above argument is valid even when $\mathfrak{S}'_{J_1}$ and $\mathfrak{S}'_{J_2}$ are the same, say, $= \mathfrak{S}'_J$.

The automorphism G of $\mathfrak{A}$ extending the isomorphism between $\mathfrak{S}'_{J_1}$ and $\mathfrak{S}'_{J_2}$ is unique. For by Lemma 1 we have $b = \sum \beta_{i_1 \cdots i_r} a_{i_1}^{(1)} \cdots a_{i_r}^{(1)}$ for any b in $\mathfrak{A}$ and since G is an automorphism $B^G = \sum \beta_{i_1 \cdots i_r} a_{i_1}^{(1)G} \cdots a_{i_r}^{(1)G} = \sum \beta_{i_1 \cdots i_r} a_{i_1}^{(2)} \cdots a_{i_r}^{(2)}$. If $J_1 = J_2 = J$ we have seen that G commutes with J. Thus the group of automorphism $\mathfrak{A}_J$ of $\mathfrak{S}'_J$ is isomorphic to the subgroup of the automorphism group of $\mathfrak{A}$, consisting of those automorphisms commutative with J. The elements of $\mathfrak{A}_J$ which correspond to inner automorphisms of $\mathfrak{A}$ form an invariant subgroup $\mathfrak{A}_J^0$ of index 1 or 2 in $\mathfrak{A}_J$. If $G \in \mathfrak{A}_J^0$ then $a^G = g^{-1}ag$ where $g \in \mathfrak{G}_J$ the group of J-orthogonal elements. The correspondence $g \to G$ is a homomorphism between $\mathfrak{G}_J$ and $\mathfrak{A}_J^0$ and it is easily seen that the elements of $\mathfrak{G}_J$ mapped into the identity of $\mathfrak{A}_J^0$ i.e. such that $g^{-1}ag = a$ for all a have the form $\delta 1$, $\delta \in$ P. If we denote the subgroup ofthese elements by $\mathfrak{D}$ we obtain the isomorphism $\mathfrak{A}_J^0 = \mathfrak{G}_J/\mathfrak{D}$.

4. Now suppose that $\mathfrak{L}$ is a normal simple Lie algebra over Φ of type A i.e. $\mathfrak{L}_\Omega \cong \Omega_n'$. Then $\mathfrak{L}$ has a representation by matrices in Ω_n' such that the linear combinations in Ω of these matrices constitute Ω_n'. Let $a_1, \cdots, a_m$ $(m = n^2 - 1)$ be a basis of $\mathfrak{L}$ over Φ, $a_j \to A_j$ in our representation, and $\mathfrak{A}$ denote the enveloping algebra over Φ of the matrices of $\mathfrak{L}$. The A_j have coördinates in an algebraic field Σ of finite degree over Φ. Hence the coördinates of all the matrices of $\mathfrak{A}$ are in Σ. Since $A_1, \cdots, A_m$ are linearly independent over Σ every element of Σ_n' is a Σ-linear combination of $A_1, \cdots, A_m$ and so $\mathfrak{L}_\Sigma \cong \Sigma_n'$. By passing to an algebraic extension we may suppose that Σ is a Galois field of finite degree over Φ with $\mathfrak{g} = (i, s, t, \cdots)$ as Galois group. If $M = (\mu_{ij})$ is any element of Σ_n we denote (μ_{ij}^s) by M^s and note that $M \to M^s$ is an automorphism of Σ_n regarded as an algebra over Φ and $(M^s)' = (M')^s$. Thus if $[a_i, a_j] = \sum \gamma_{ijk} a_k$ $(\gamma \in \Phi)$ then $[A_i^s, A_j^s] = \sum \gamma_{ijk} A_k^s$ for all s in $\mathfrak{g}$ and the correspondence $M = \sum \mu_j A_j \to M^S = \sum \mu_j A_j^s$ is an automorphism S of Σ_n' over Σ. By Lemma 3 S is either inner i.e. there exists a G_s such that $A_j^S = A_j^s = G_s^{-1} A_j G_s$ or else there exists a G_s such that $A_j^s = -G_s^{-1} A_j' G_s$. As shown above, if $n = 2$ the automorphisms of Σ_n' are all inner but if $n > 2$ the automorphisms of the form $A \to -GA'G$ are outer. Let G_s be a fixed matrix satisfying our conditions. Then any other matrix H_s having this property is a multiple of G_s. For $G_s^{-1}H_s$ commutes with all A_j and hence with all the elements of Σ_n. It follows that $G_s^{-1}H_s = \rho_s 1$, $H_s = \rho_s G_s$.

Let $\mathfrak{g}_0$ be the totality of elements s of $\mathfrak{g}$ such that the corresponding S is inner. Then the following table holds:

1. If $s, t \in \mathfrak{g}_0$ then $st \in \mathfrak{g}_0$ and $G_{st} = \rho_{s,t} G_t G_s^t$, $\rho_{s,t} \neq 0$ in Σ.
2. If $s \in \mathfrak{g}_0$, $t \notin \mathfrak{g}_0$ then $st \notin \mathfrak{g}_0$ and $G_{st} = \rho_{s,t} G_t G_s^t$.
3. If $s \notin \mathfrak{g}_0$, $t \in \mathfrak{g}_0$ then $st \notin \mathfrak{g}_0$ and $G_{st} = \rho_{s,t} (G_t')^{-1} G_s^t$.
4. If $s \notin \mathfrak{g}_0$, $t \notin \mathfrak{g}_0$ then $st \in \mathfrak{g}_0$ and $G_{st} = \rho_{s,t} (G_t')^{-1} G_s^t$.

We shall prove 1 and leave the others, which are derived in a similar fashion, to the reader. Assume $A_j^s = G_s^{-1} A_j G_s$ and $A_j^t = G_t^{-1} A_j G_t$. Then $A_j^{st} = (A_j^s)^t = (G_s^t)^{-1} A_j^t G_s^t = (G_s^t)^{-1} G_t^{-1} A_j G_t G_s^t$. Hence $st \in \mathfrak{g}_0$ and G_{st} is a multiple of $G_t G_s^t$. Our table shows that $\mathfrak{g}_0$ is an invariant subgroup of index 1 or 2 in $\mathfrak{g}$.

Suppose first that $\mathfrak{g}_0 = \mathfrak{g}$. Set $A_0 = 1$ and $A = \sum_{i=0}^m \alpha_i A_i$ where $\alpha_i \in \Phi$. Then

$$A^s = G_s^{-1} A G_s \qquad s \in \mathfrak{g}$$

The matrices satisfying these equations form an algebra $\mathfrak{B}$ over Φ containing the algebra $\mathfrak{A}$. On the other hand if $N = \sum \nu_i A_i$ is in $\mathfrak{B}$ then

$$\sum \nu_i^s A_i^s = N^s = G_s^{-1} N G_s = \sum \nu_i A_i^s$$

and so $\nu_i^s = \nu_i \in \Phi$. Thus $\mathfrak{B}$ has order $m + 1 = n^2$ and $\mathfrak{B} \supset \mathfrak{A} \supset$ the matrices $\sum_{j=1}^m \alpha_j A_j$ of $\mathfrak{L}$. Hence either $\mathfrak{A} = \mathfrak{B}$ or $\mathfrak{A}$ is the set of matrices of $\mathfrak{L}$. But the latter is impossible since the matrices of trace 0 do not form an algebra under ordinary multiplication and so $\mathfrak{A} = \mathfrak{B}$. $\mathfrak{A}' \supset$ the matrices $\sum \alpha_j A_j$ of $\mathfrak{L}' = \mathfrak{L}$ and since tr $A_0 = n$, $A_0 \notin \mathfrak{A}'$ and $\mathfrak{A}'$ is the totality of matrices

$\sum \alpha_j A_j$ i.e. $\mathfrak{A}' \cong \mathfrak{L}$. Since $\mathfrak{A}_\Sigma \cong \Sigma_n$, $\mathfrak{A}$ is normal simple with Σ as a splitting field. It follows that the traces of the elements of $\mathfrak{A}$ and the coefficients of the characteristic equations of the matrices of $\mathfrak{L}$ are all in Φ and hence $\mathfrak{L}$ has type A_{I} in the sense of Landherr.[20]

If $\mathfrak{L}$ is of type A_{II} $\mathfrak{g}/\mathfrak{g}_0$ has order 2 and $n > 2$. Corresponding to $\mathfrak{g}_0$ there is a quadratic subfield $\mathrm{P} = \Phi(q)$ such that the group of Σ over P is $\mathfrak{g}_0$. $\mathfrak{L}_{\mathrm{P}}$ is isomorphic to the set of matrices $\sum_{j=1}^m \xi_j A_j$, $\xi \in \mathrm{P}$ and $(\mathfrak{L}_\rho)_\Sigma \cong \Sigma_n'$. Since the automorphisms s of the Galois group of Σ over P induce inner automorphisms S in Σ_n' we conclude by the above argument $\mathfrak{L}_{\mathrm{P}} \cong \mathfrak{B}'$ where $\mathfrak{B}$ is a normal simple algebra over P, namely, the set of matrices $\sum_{i=0}^m \xi_i A_i$, $\xi \in \mathrm{P}$. Now let t be an element of $\mathfrak{g}$ not in $\mathfrak{g}_0$. Then t induces an automorphism distinct from the identity in P and so if $\xi \in \mathrm{P}$, $\xi^t = \bar{\xi}$. Since $t \notin \mathfrak{g}_0$, $A_j^t = -G_t^{-1} A_j' G_t$ and hence for $H = G_t'$ we have $-A_j = H^{-1}(A_j')^t H$. Let $A_0 = q1$. Then the correspondence $X = \sum \xi_i A_i \rightarrow X^J = H^{-1}(X')^t H = -\sum \bar{\xi}_i A_i$ is readily seen to be an i.a.a. of second kind in $\mathfrak{B}$ regarded as an algebra over Φ. $\mathfrak{S}_J$ is the set of matrices $\sum \alpha_i A_i$, $\alpha \in \Phi$ and $\mathfrak{S}_J'$ whose order is m is the set $\sum_1^m \alpha_j A_j$ i.e. $\mathfrak{S}_j' \cong \mathfrak{L}$. Hence by Lemma 1 $\mathfrak{B} = \mathfrak{A}$, the enveloping algebra of $\mathfrak{S}_J'$ the set of matrices of $\mathfrak{L}$.

THEOREM 4. *If $\mathfrak{L}$ is a Lie algebra of type A_{II} then $n > 2$ and $\mathfrak{L} \cong \mathfrak{S}_J'$ where J is an i.a.a. of second kind in a simple algebra $\mathfrak{A}$.*

5. In this section we consider the question of cogredience of i.a.a.'s in $\mathfrak{A}$.

By Wedderburn's theorem $\mathfrak{A} \cong \mathfrak{F}_r$ the r-rowed matrix algebra with coördinates in a division algebra, and as has been shown by Albert[21] there is an i.a.a. $a \rightarrow \bar{a}$ in $\mathfrak{F}$ such that $J_0 : A = (a_{ij}) \rightarrow (\bar{a}_{ji}) = \bar{A}' = A^{J_0}$ $(a_{ij} \in \mathfrak{F})$ is an i.a.a. of second kind in $\mathfrak{A}$. Since $J_0^{-1}J = P$ is an automorphism leaving the elements of P invariant it is inner, say $A^P = P^{-1}AP$ where the matrix P is determined only to within a multiple $\neq 0$ in P. Thus $J = J_0 P$ and $A^J = P^{-1}\bar{A}'P$. The condition that J is involutorial is $A^{J^2} = P^{-1}\bar{P}'A(\bar{P}')^{-1}P = A$ i.e. $\bar{P}' = \rho P$, $\rho \neq 0$ in P. But then $P = \bar{\rho}\bar{P}' = \bar{\rho}\rho P$ so that $\bar{\rho}\rho = 1$. Hence there exists a σ in P such that $\rho = \sigma(\bar{\sigma})^{-1}$ and σP is hermitian.[22] We may therefore suppose that P is hermitian in the expression $P^{-1}\bar{A}'P$ for A^J. P will then be determined to within a multiple $\neq 0$ in Φ. Thus associated with the i.a.a. we have a unique ray $\{P\}$ of non-singular hermitian matrices consisting of all Φ-multiples $\neq 0$ of a fixed one of the set.

We shall say that the rays $\{P\}$ and $\{Q\}$ are *cogredient* if for any $P \in \{P\}$ and $Q \in \{Q\}$ here is an S in $\mathfrak{F}_r$ and a μ in Φ such that $Q = \mu\bar{S}'PS$ i.e. Q is cogredient in the usual sense to a multiple of P.

If $K = S^{-1}JS$ where $A^S = S^{-1}AS$ and $\{Q\}$ is the ray of K then $A^K = Q^{-1}\bar{A}'Q = (\bar{S}'PS)^{-1}\bar{A}'(\bar{S}'PS)$ and hence $\mu\bar{S}'PS = Q$ or $\{P\}$ and $\{Q\}$ are cogredient. The converse of this is immediate. If there exist outer automor-

[20] L. p. 50.

[21] Albert loc cit. in (4) p. 897.

[22] Cf. Albert p. 897.

phisms of $\mathfrak{A}$ and S_0 is a fixed one then any other outer automorphism is the product of S_0 by an inner automorphism. The mapping $S_0^{-1}J_0S_0$ is an i.a.a. of second kind which we may suppose has the form $A \to U_0^{-1}\bar{A}'U_0$ where U_0 is a fixed hermitian matrix in the ray $\{U_0\}$. Consider $S_0^{-1}JS_0 = S_0^{-1}J_0PS_0 = (S_0^{-1}J_0S_0)(S_0^{-1}PS_0)$. As we have seen $S_0^{-1}PS_0$ is the inner automorphism $A \to (P^{S_0})^{-1}AP^{S_0}$. Hence $A^{S_0^{-1}JS_0} = (P^{S_0})^{-1}U_0^{-1}\bar{A}'U_0P^{S_0}$. Thus the $\{U_0P^{S_0}\}$ determines an i.a.a. cogredient to J also and the ray of any i.a.a. cogredient to J is cogredient either to $\{P\}$ or to $\{U_0P^{S_0}\}$.

In particular if $\mathfrak{F} = \mathrm{P}$ $(r = n)$ we define J_0 as before and let S_0 be the automorphism $A \to \bar{A}$. Then $S_0^{-1}J_0S_0 = J_0$ and we may suppose that $U_0 = 1$. Hence the i.a.a. cogredient to J have rays cogredient to either $\{P\}$ or $\{\bar{P}\}$. However we note that P and $\bar{P}$ are cogredient matrices. For P is cogredient to a diagonal matrix D, say $P = \bar{V}'DV$ and since D is hermitian $\bar{D} = D$. Then $\bar{P} = V'D\bar{V}$ and hence $\bar{P} = V'(\bar{V}')^{-1}PV^{-1}\bar{V} = \bar{W}'PW$ where $W = V^{-1}\bar{V}$. It follows in this case that a necessary and sufficient condition that J and K be cogredient is that $\{P\}$ and $\{Q\}$ be cogredient rays.

UNIVERSITY OF CHICAGO.

ANNALS OF MATHEMATICS
Vol. 39, No. 1, January, 1938

ON THE TOPOLOGY OF REAL PLANE ALGEBRAIC CURVES

BY I. PETROWSKY

(Received February 4, 1937)

INTRODUCTION

As early as 1876 Harnack[1] showed that the maximal number of components (maximal connected subsets) of a real algebraic curve of order n in the projective plane is precisely $\frac{1}{2}(n-1)(n-2)+1$. At the same time Harnack proposed a process for the construction of curves with this maximal number of components. Such curves we shall call in the sequel, M-curves. Harnack showed that these M-curves have no singular points.

Take a sphere in the three-dimensional space in which the projective plane containing our algebraic curve is situated, and join the centre of this sphere to every point of the projective plane by a straight line. We thus project the plane on the sphere. A component of an algebraic curve is called an "oval" (or an "even" component) if its projection on the sphere consists of two ordinary closed curves. If this projection on the sphere S consists of a single closed curve the corresponding component is called "odd." Algebraic curves having no real[2] singular points possess at most one odd component. Hence every algebraic curve (having no real singular points) of even order consists of ovals only while a curve of odd order has (besides ovals) exactly one odd component.

In 1891 D. Hilbert[3] proposed a new method of constructing M-curves. In the same work Hilbert announced without proof that an M-curve of order 6 cannot have all its ovals lying outside each other. At least one of these ovals must lie within another oval. Here the words "an oval lies within another oval" mean that the cone projecting the first oval on S lies within the projecting cone of the second oval. Hilbert considers this a remarkable fact, since it proves that M-curves cannot have a too simple topological structure. In his report to the International Mathematical Congress in 1900 on modern problems of mathematics Hilbert considers the investigation of the topology of M-curves and of the corresponding algebraic surfaces as most timely.[4] After a series of attempts the above mentioned theorem announced by Hilbert was at last proved in 1911 by K. Rohn.[5] In the same work Rohn proved that an

[1] Math. Annalen, Bd. X. p. 189.

[2] In the sequel, whenever we mention the singularities of the curve we shall mean only *real* singular points.

[3] Math. Annalen, Bd. XXXVIII. p. 115–138 (1891).

[4] Göttingen Nachrichten, 1901.

[5] K. Rohn, Leipzig. Ber. Dezember 1911. *Die Ebene Kurven 6. Ordnung mit elf Ovalen.*

algebraic curve of the sixth order cannot possess an oval with ten other ovals of the same curve interior to it.

The object of the present paper is to give a general method which enables us to obtain some of the above mentioned results of Rohn and to extend them to plane curves of arbitrary order. We shall obtain the following results: *When n is even, an algebraic curve of order n consists of at most $\frac{1}{8}(3n^2 - 6n) + 1$ ovals exterior to each other. Curves having this number of ovals exterior to each other do exist.* They may be constructed by the same method as that Harnack used for M-curves.

Hilbert's construction of M-curves leads us to curves having among their components $[\frac{1}{2}n - 1]$ (where $[k]$ denotes the greatest integer contained in k) consecutive ovals lying within each other (i.e. every oval within the preceding oval). In order to extend our theorem to such curves we shall call an oval O of a curve of even order, $F(x, y) = 0$, "*positive*" ("*negative*") if when we cross the oval in the outward direction the value of the function $F(x, y)$ decreases (resp. increases). In the case of an even n we can always suppose (changing the sign of $F(x, y)$ if necessary) that the ovals not lying within other ovals or lying inside an even number of consecutive ovals are positive while the ovals lying within an odd number of ovals are negative. Then we can prove the following general theorem which is an extension of the one formulated above.

Denote by p the number of positive ovals and by m the number of negative ovals of an algebraic curve of an even order n. Then

$$|p - m| \leqq \frac{3n^2 - 6n}{8} + 1$$

and there are curves for which this limit is reached.

The proof will give us a somewhat more precise bound for $|p - m|$. We shall see, in particular that the number of ovals lying outside each other and all within the same oval can not exceed $\frac{1}{8}(3n^2 - 6n) + 1$, if the algebraic curve does not contain other ovals. We can prove by examples that this number can not exceed the precise limit by more than 3 in the case when $n = 4k$ (k an integer) and by more than one in the case when $n = 4k + 2$. Rohn's result (that there do not exist curves of order 6 with ten ovals all lying inside the eleventh oval) shows that this bound is not precise.

In the case of an odd order n, Harnack's process already yields the M-curve consisting (besides an odd branch) of ovals none of which lies inside another oval. This case may be treated with the same methods which we used in the case of an even n with the following modifications.

Just as in the even case, any finite oval of the curve of odd order $F(x, y) = 0$ shall be called *positive* if when crossing the oval from the inside we pass from the values of $F(x, y) > 0$ to the values of $F(x, y) < 0$. In the opposite case we shall call the oval *negative*. This definition evidently does not apply to the oval of a curve of odd order intersecting the line at infinity. These ovals we shall call "*zero-ovals.*"

Now *for an odd n we can prove that*

$$\left| m - p - \Delta + \frac{k+1}{2} \right| \leqq \frac{3n^2 - 4n + 1}{8},$$

where m is the number of negative ovals,
p " " " " positive ovals,
k " " " " real points at infinity of the curve $F(x, y) = 0$ *and*
Δ is a certain positive integer ≦ k, depending upon the character of the intersection of $F(x, y) = 0$ *with the line at infinity.* A more precise definition of this number will be given in §2 (see lemma 3 and the proof of the second Fundamental Theorem).[6]

We shall give in this paper examples showing that *this is the precise limit of* $|p - m|$, *in the sense that the limit is reached by some curves.*

As an immediate corollary of this theorem we obtain the following result. The odd branch of an algebraic curve of an odd order n together with the line at infinity divide the projective plane in several regions. If the algebraic curve $F(x, y) = 0$ contains only ovals lying in one of the above regions and which are exterior to each other, then their number is not more than

$$\frac{3n^2 - 4n + 1}{8} + \left| \Delta - \frac{k+1}{2} \right| \leqq \frac{3n^2 - 4n + 1}{8} + \frac{k-1}{2}.$$

Since the line at infinity does not differ (in the sense of projective geometry) from any other straight line in the plane we can substitute in this theorem any other straight line for the line at infinity.

The proofs of all these theorems are based on a formula of Jacobi-Euler concerning solutions of systems of algebraic equations[7] and on the consideration of the deformations of lines $F(x, y) = C$ when C crosses the critical values of $F(x, y)$. These last investigations are analogous to those of Morse[8] on the critical points of a function.

The method which we apply to the investigation of plane algebraic curves admits of a natural generalization to algebraic hypersurfaces

$$F(x_1, \cdots, x_d) = 0$$

in a projective space of an arbitrary dimension d. This question we propose to treat extensively in another paper.

The results of the present paper could be obtained as a corollary from the theorems concerning hypersurfaces in a d-dimensional space (which shall be considered in the subsequent paper). Nevertheless we preferred to treat them

---

[6] I failed to take into account the number Δ in my note in Comptes Rendus, Paris, 1933, t. 197, p. 1270.

[7] Euler, Instit. Calc. Integr. Petrop. 1768–70, 2, 1169. Jacobi, Gesammelte Werke, Bd. 3, p. 329–354. Cf. also Kronecker's Werke, Bd. I, p. 133.

[8] Transactions of the American Math. Society, V. 27, 1225, p. 345.

separately because they are quite evident geometrically and besides they help to understand the character of the general method of our investigation.

We recall that everywhere in the sequel we consider only algebraic curves without (real) singularities.

1. Some Preliminary Lemmas

Lemma 1. *Let*

$$F(x, y) = 0 \tag{1}$$

be the equation of a real[9] *algebraic curve. If we now continuously vary its coefficients the topological structure of the curve changes only when the coefficients pass through the values for which the curve has a singularity, i.e. through values taken at points which satisfy simultaneously* (1) *and*

$$F'_x(x, y) = 0; \qquad F'_y(x, y) = 0. \tag{2}$$

Proof. Evidently the topological structure of the curve (1) can change only when two different points of this curve unite, or when an isolated point appears or disappears, i.e. when the curve has singular points. Of course these singularities may lie at infinity. Then we must pass to homogeneous coordinates in order to take care of these points.

Lemma 2. *Let*

$$F(x, y) = 0 \tag{3}$$

be the equation of a (plane) algebraic curve of order n without singularities. Then we can so change its equation without changing either its order or its topological structure that the equations (2) *will have* $(n - 1)^2$ *different finite solutions, real or imaginary, and that for any two different real solutions* (x_1, y_1) *and* (x_2, y_2) *of* (2) *we have*

$$F(x_1, y_1) \neq F(x_2, y_2).$$

Proof. By lemma 1 if we vary the coefficients of the equation (3) so little that no singularities arise the curve does not change its topological structure (if only as we have supposed, the curve (3) has no singular points). On the other hand the conditions that the system (2) have infinite solutions or equal solutions or an infinity of solutions etc. are that certain polynomials in its coefficients vanish. Therefore if only these polynomials do not vanish identically we can always vary the coefficients a little in such a way that the modified equation satisfies the conditions of lemma 2. The fact that the polynomials spoken of above do not vanish identically can be easily established by constructing examples of functions satisfying the conditions of lemma 2.

---

[9] In future we shall consider only the equations of algebraic curves with real coefficients.

LEMMA 3. *Let (x_0, y_0) be a real finite critical point of the function $F(x, y)$,* i.e. a point satisfying (2). *If at this point*

$$D(x_0, y_0) = \begin{vmatrix} \dfrac{\partial^2 F}{\partial x^2} & \dfrac{\partial^2 F}{\partial x \partial y} \\ \dfrac{\partial^2 F}{\partial x \partial y} & \dfrac{\partial^2 F}{\partial y^2} \end{vmatrix} > 0$$

(such critical points we shall call *plus-points*) *then when C varies from $F(x_0, y_0) + \epsilon$ to $F(x_0, y_0) - \epsilon$ $(\epsilon > 0)$ the difference $p - m$ between the number p of positive ovals of the curve $F(x, y) = C$ and the number m of negative ovals*[10] *of this curve increases by* 1.

If $D(x_0, y_0) < 0$ (such critical points we shall call *minus-points*) *and if the polynomial $F(x, y)$ is of even order, then when C varies from $F(x_0, y_0) + \epsilon$ to $F(x_0, y_0) - \epsilon$ the difference $p - m$ decreases always by* 1, *except once when it decreases by* 2.

Finally if $D(x_0, y_0) < 0$ and if $F(x, y)$ is of odd order, then the difference $p - m$ either decreases by one or remains fixed.

It is supposed here that all critical values (i.e. values at critical points) of $F(x, y)$ are different and that ϵ is so small that the interval

$$(F(x_0, y_0) - \epsilon, \qquad F(x_0, y_0) + \epsilon)$$

contains no critical values of $F(x, y)$ except $F(x_0, y_0)$.

PROOF. If $D(x_0, y_0) > 0$ then at the critical point (x_0, y_0) we have a maximum or a minimum of $F(x, y)$. If $F(x_0, y_0)$ is a maximum, then as C in decreasing passes through the critical value, *a new positive oval arises.* If $F(x_0, y_0)$ is a minimum value of $F(x, y)$ then *a negative oval disappears* when C passes through the value $F(x_0, y_0)$.

If $D(x_0, y_0) < 0$ we have at the point (x_0, y_0) a saddle point of the function $F(x, y)$. We must distinguish here between the cases of n even or odd.

If the order n of $F(x, y)$ is even, there are three possibilities:

1. A positive oval touches another (positive or negative) oval. In this case *a positive oval disappears.*

2. An oval (positive or negative) touches itself. Here we must distinguish two cases.

2a. An "odd branch" can be traced in the region where $F(x, y) > F(x_0, y_0) + \epsilon$ or in the region where $F(x, y) < F(x_0, y_0) - \epsilon$ $(\epsilon > 0)$. Then, as C varies from $F(x_0, y_0) + \epsilon$ to $F(x_0, y_0) - \epsilon$ *a new negative oval appears.*

2b. Though we can trace odd branches in the regions $F(x, y) > F(x_0, y_0) - \epsilon$ and $F(x, y) < F(x_0, y_0) + \epsilon$, no odd branches can be traced in either of the regions $F(x, y) < F(x_0, y_0) - \epsilon$, $F(x, y) > F(x_0, y_0) + \epsilon$. It is evident that this case may present itself *at most once* as C varies from $+\infty$ to $-\infty$. When

[10] Cf. introduction.

it does, as C passes from $F(x_0, y_0) + \epsilon$ to $F(x_0, y_0) - \epsilon$, *a positive oval becomes negative.*

The cases 2a and 2b resp. can be defined also as the cases when the sign of the outer ovals (i.e. of ovals not contained in other ovals of the curve) does not change, resp. changes from $+$ to $-$.

Two negative ovals cannot touch when C decreases because then such ovals recede from each other.

If n is odd (and $D(x_0, y_0) < 0$) we can distinguish the following cases:

1. A positive oval touches another oval or the odd branch. Then this *positive oval disappears.*

2. A (positive or negative) oval touches itself. Then *a new negative oval arises.*

3. A zero-oval or an odd branch touches itself or another zero-oval. In order to represent in the simplest manner all the possibilities which can arise in this case we shall consider the projection of our projective plane upon a hemisphere S bounded by a great circle L representing the line at infinity. Consider the set of all points of this hemisphere which correspond to the points of the plane where $F(x, y) > C$. This (open) set contains a number of regions $G_1, G_2, \cdots, G_{k_1}$, the boundaries of which contain segments $l_1, l_2, \cdots, l_{k_2}$ of L. When C decreases these regions expand while the segments $l_1, l_2, \cdots, l_{k_2}$ remain intact. If now a boundary of one of these regions G_i touches itself *a new negative oval arises*; if it touches the boundary of another such region then *the regions coalesce and a zero-oval disappears or a new zero-oval arises* (because this oval intersects the line at infinity). In this case the number of the regions G_i decreases by one. Thus in all three cases lemma 3 is proved.

LEMMA 4. *Suppose that the curve $F(x, y) = C$ meets the line at infinity in k different points (k evidently does not depend on C) and that all critical points*[11] *of $F(x, y)$ are finite and different; then the difference D between the number of minus-points and the number of plus-points of this function is $k - 1$.*

PROOF. Denote by C_M the maximal critical value of $F(x, y)$ and C_m the minimal critical value and consider again the projection of the plane (x, y) upon the hemisphere S bound by the great circle corresponding to the line at infinity. Let M_C be the set of all points of the hemisphere which correspond to the points (x, y) of the plane for which $F(x, y) > C$.

If $k \geqq 1$ then when $C > C_M$ the set M_C consists of k regions $G_1, \cdots, G_k$, k being the number of (real) points in which the curves meet the line at infinity; and when $C < C_m$ the set M_C consists of a single simply-connected region; Hence when we vary C from $C > C_M$ to $C < C_m$ all k regions G_i must coalesce. This would give exactly $k - 1$ minus-points because two of the regions G_i can coalesce only when C passes through a minus-point. Besides new ovals may arise and subsequently disappear while C varies from $C_M + \epsilon$ to $C_m - \epsilon$. But as every new positive oval can arise only in a plus-point and disappear in a

[11] The number of which is finite.

minus-point, while a negative oval arises in a minus-point and disappears in a plus-point; and as in every plus-point a positive oval arises or a negative oval disappears; and in a minus-point either a negative oval arises, or a positive oval disappears, or two of the original regions G_i coalesce; all this does not affect the difference D which remains $k - 1$.

If $k = 0$ then either when $C > C_M$ the curve consists of a single negative oval and when $C < C_m$ the curve is imaginary, containing no real points of the plane, or conversely when $C > C_M$ the curve is imaginary and when $C < C_m$ the curve consists of a single positive oval. In both cases there must exist a plus-point in which this oval arises or disappears. All other ovals, arising and subsequently disappearing as C varies from $C_M + \epsilon$ to $C_m - \epsilon$, give rise to an equal number of plus and minus points. Therefore the difference D is -1 and the lemma remains true in this case also.

LEMMA 5. *Let $F_1(x, y)$ and $F_2(x, y)$ be two polynomials of degree n in x and y vanishing simultaneously at exactly n^2 different (finite) points, and $f(x, y)$ a polynomial of degree $l < n$ in x, y not identically zero. Then $f(x, y)$ cannot vanish in more than nl points at which $F_1(x, y) = 0$ and $F_2(x, y) = 0$ simultaneously.*

PROOF. The polynomials F_1 and F_2 have no common factor because they vanish simultaneously at a discrete set of points. Denote by $M_1(x, y)$ the greatest common factor of F_1 and f, and by $M_2(x, y)$ the greatest common factor of F_2 and f and let the degrees of $M_1(x, y)$ and $M_2(x, y)$ be n_1 and n_2 (n_1 and n_2 can of course be equal to 0). F_1 and F_2 having no common factor, we can write

$$F_1(x, y) = M_1(x, y) \cdot \bar{M}_1(x, y),$$

$$F_2(x, y) = M_2(x, y) \cdot \bar{M}_2(x, y),$$

$$f(x, y) = M_1(x, y) \cdot M_2(x, y) \cdot M(x, y),$$

where $\bar{M}_1(x, y)$, $\bar{M}_2(x, y)$ and $M(x, y)$ are polynomials in x, y.

The functions F_1, F_2 and f can vanish simultaneously only at those (finite) points of the plane (x, y) which satisfy at least one of the following three systems:

1. $M_1(x, y) = 0, F_2(x, y) = 0,$
2. $M_2(x, y) = 0, F_1(x, y) = 0$
3. $M(x, y) = 0, F_1(x, y) = 0$ (or $F_2(x, y) = 0$).

The left-hand members of each of these systems of equations are relatively prime. Therefore the first of these systems has at most $n_1 n$ (finite) solutions, the second at most $n_2 n$, and the third at most $(l - n_1 - n_2) \cdot n$.[12] The sum of these numbers is $ln < n^2$ because $l < n$. Therefore the lemma is proved.

LEMMA 6. *Let A be a fixed imaginary number different from 0 and $f(x, y)$ a polynomial in x, y with real coefficients. Then the condition that the real part of the product $A[f(x, y)]^2 = 0$ at a given point (x_0, y_0)* (real or imaginary) *can be*

[12] Cf. Encyklopädie der Math. Wissenschaften, Bd. I, p. 263, Kronecker'sche Methode.

expressed in the form of a linear homogeneous equation, with real coefficients, in the coefficients of $f(x, y)$.

Proof. Let $A = a + bi$ and $f(x_0, y_0) = c + di$, where the numbers a, b, c, d are real. Then

$$A[f(x_0, y_0)]^2 = (a + bi)(c + di)^2$$

and

$$R\{A[f(x_0, y_0)]^2\} = ac^2 - ad^2 - 2bcd = c^2(a - 2bx - ax^2)$$

where $x = d/c$ and $R\{k\}$ denotes the real part of k. The equation

$$ax^2 + 2bx - a = 0$$

has (for any real numbers a, b) at least one real finite root. Denote this root by k. Then $d = kc$ implies $R\{A[f(x, y)]^2\} = 0$ and the equation $d = kc$ is a linear homogeneous equation with real coefficients in the coefficients of $f(x, y)$, q.e.d.

2. Two Fundamental Theorems

First Fundamental Theorem. *Denote by* p *the number of positive ovals of a curve* $F(x, y) = 0$ *of even order* n *and by* m *the number of its negative ovals. Then*

$$-\frac{3n^2 - 6n}{8} - \delta \leqq p - m \leqq \frac{3n^2 - 6n}{8} + 1 - \delta,$$

where $\delta = 0$ *if the outer ovals are positive and* $= 1$ *if the outer ovals are negative.*

From this formula it follows of course that

$$|p - m| \leqq \frac{3n^2 - 6n}{8} + 1.$$

Proof. The critical points of $F(x, y)$ are given by (2). By Lemma 2 we may always suppose that this system possesses $(n - 1)^2$ different finite solutions and that if (x_1, y_1) and (x_2, y_2) are two different solutions then $F(x_1, y_1) \neq F(x_2, y_2)$. In this case the following theorem of Euler-Jacobi holds:

$$\sum_{i=1}^{(n-1)^2} \frac{P(x_i, y_i)}{J(x_i, y_i)} = 0 \tag{4}$$

where $J(x, y)$ is the Jacobian of (2) and $P(x, y)$ is an arbitrary polynomial in x and y of degree lower than $2n - 4$. The sum in the left hand member of (4) is taken over all solutions of the system (2).[13] The system (2) having no multiple solutions, the denominator $J(x_i, y_i)$ of any member of the sum in (4) is not zero.

In particular (4) holds if we take

$$P(x, y) = F_1(x, y)[f(x, y)]^2 \tag{5}$$

[13] Jacobi, *Werke* Bd. 3, p. 329. Cf. *Kroneckers Werke* Bd. 1, p. 133.

where $f(x, y)$ is an (arbitrary) polynomial of degree $\frac{1}{2}(n - 4)$ and

$$F_1(x, y) = nF(x, y) - xF'_x(x, y) - yF'_y(x, y).$$

The function $F_1(x, y)$ thus defined is of degree $n - 1$ at most (this follows easily from Euler's theorem on homogeneous functions); at any critical point of the function F, F_1 has the same sign as F. The function $f(x, y)$ may be so chosen that it has real coefficients and vanishes in $\dfrac{\left(\frac{n-4}{2}+1\right)\left(\frac{n-4}{2}+2\right)}{2} - 1$ arbitrarily chosen critical points of $F(x, y)$. Lemma 5 shows that it cannot vanish in all critical points of $F(x, y)$.

After these preliminary remarks we pass on to the proof of our first fundamental theorem. Denote by k the number of points in which the curve $F(x, y) = C$ meets the line at infinity. If $k > 0$ and $C > C_M$ the curve consists of $\frac{1}{2}k$ positive ovals. When C decreases from $C_M + \epsilon$ to 0 the curve obtains

$$p + \alpha - \frac{k}{2} \quad \text{new positive ovals,}$$

$$m + \beta \quad \text{new negative ovals,}$$

and loses

$$\alpha \quad \text{positive ovals,}$$

$$\beta \quad \text{negative ovals.}$$

Accordingly C passes through the critical values of $F(x, y)$ in

$$p + \alpha + \beta - \frac{k}{2} \quad \text{plus-points (A-points) and}$$

$$m + \alpha + \beta - \delta \quad \text{minus-points (B-points)}$$

(see lemma 3). Here $\delta = 0$ if the outer ovals are positive and $= 1$ if they are negative. These last formulae, giving the number of A and B points which were proved under the assumption that $k > 0$, still hold when $k = 0$. This can be proved by direct consideration of the two possible cases: when we have at the outset ($C > C_M$), 1) an imaginary curve ($\delta = 0$), or 2) a negative oval ($\delta = 1$).

We have thus proved that $F(x, y) > 0$ at

$$p + \alpha + \beta - \frac{k}{2} \quad \text{plus-points and at}$$

$$m + \alpha + \beta - \delta \quad \text{minus-points of } F.$$

But by lemma 4 there exist besides these

$$\frac{(n-1)^2 - 2\gamma + 1}{2} - p - \alpha - \beta \quad \text{plus-points } (B'\text{-points}) \text{ and}$$

$$\frac{(n-1)^2 - 2\gamma - 1}{2} - m - \alpha - \beta + \frac{k}{2} + \delta \quad \text{minus-points } (A'\text{-points})$$

where $F(x, y) < 0$. Here 2γ denotes the number of imaginary critical points of $F(x, y)$.

At the A and A'-points, $F/J > 0$. There are all told

$$\frac{(n-1)^2 - 1}{2} - \gamma + p - m + \delta$$

such points. At the B- and B'-points, $F/J < 0$. The number of these points is

$$\frac{(n-1)^2 + 1}{2} - \gamma - p + m - \delta.$$

On the other hand it is easy to see that the sum of the number of A-points and A'-points exceeds $\frac{1}{8}n(n-2) - \gamma - 1$. In fact, supposing the contrary, chose $\frac{1}{8}n(n-2)$ coefficients of the polynomial $f(x, y)$ entering in (4) and (5) so that it vanishes at all A- and A'-points and so that, besides, the real parts of the components $P(x_i, y_i)/J(x_i, y_i)$ of the sum entering in (4) corresponding to imaginary solutions of the system (2) vanish also (we suppose that $f(x, y)$ does not vanish identically). This last condition can be fulfilled in consequence of lemma 6. $f(x, y)$ differs from 0 in at least one real critical point of $F(x, y)$. In fact if

$$\frac{n(n-2)}{8} - \gamma - 1 \geqq 0$$

(evidently only this case need be considered) then the number of the real critical points of $F(x, y)$, which is $(n-1)^2 - 2\gamma$, exceeds the number

$$(n-1)^2 - \frac{n(n-2)}{4} + 2 - 1 = \tfrac{3}{4}n^2 - \tfrac{3}{2}n + 2,$$

while the number of critical points of $F(x, y)$ where $f(x, y)$ vanishes can not exceed (by lemma 5)

$$\frac{(n-1)(n-4)}{2}.$$

Then the left-hand member of (4) reduces to a sum of non-positive numbers which cannot all vanish simultaneously, while the right hand member is 0; which is absurd. In the same way we can prove that the number of the B- and B'-points exceeds $\frac{1}{8}n(n-2) - \gamma - 1$.

If we combine these last results and those obtained above there results

$$-\frac{3n^2 - 6n}{8} \leqq p - m + \delta \leqq \frac{3n^2 - 6n}{8} + 1, \quad \text{q.e.d.}$$

In particular when $n = 6$ we obtain

$$-9 - \delta \leqq p - m \leqq 10.$$

When $m = 0$ we obtain $p \leqq 10$, i.e. an algebraic curve of sixth order can not consist of more than 10 positive ovals; in other words of more than 10 ovals lying exterior to each other (Hilbert-Rohn's Theorem).

SECOND FUNDAMENTAL THEOREM. *Denote again by p the number of positive ovals and by m the number of negative ovals of a curve of order n without singularities; and let n this time be an odd number. Then*

$$\left| m - p - \Delta + \frac{k+1}{2} \right| \leqq \frac{3n^2 - 4n + 1}{8}$$

where k is the number of real points at infinity of our curve, and Δ is the number of the regions G_i, corresponding to the curve $F(x, y) = 0$, which were defined in lemma 3 ($1 \leqq \Delta \leqq k$). We can also write this formula in the following more symmetrical form

$$\left| m - p + \frac{\Delta^- - \Delta^+}{2} \right| \leqq \frac{3n^2 - 4n + 1}{8},$$

where $\Delta^+ = \Delta$, while $\Delta^- = k + 1 - \Delta$ is the number of the regions G_i, corresponding to the curve $-F(x, y) = 0$.

PROOF. Our starting point is again the identity (4) with

$$P(x, y) = F_1(x, y)[f(x, y)]^2 \tag{6}$$

$f(x, y)$ being this time a polynomial of degree $\frac{1}{2}(n - 5)$. Suppose that the curve

$$F(x, y) = C \tag{7}$$

meets the line at infinity in k points. k is then an odd number independent of C. When $C > C_M$ the curve (7) evidently consists of its odd branch. If the curve $F(x, y) = 0$ consists (besides the odd branch) of p positive and m negative ovals, and if Δ is the number of the regions G_i (defined in lemma 3), then as C was varied from $C_M + \epsilon$ to 0

$$\begin{aligned} p + \alpha &\quad \text{new positive ovals,} \\ m + \beta &\quad \text{new negative ovals} \end{aligned}$$

appeared (where α and β are non-negative integers) and

$$\begin{aligned} \alpha &\quad \text{positive ovals,} \\ \beta &\quad \text{negative ovals,} \end{aligned}$$

were lost, while the number of the regions G_i was decreased by $k - \Delta$. Consequently there exist

$$p + \alpha + \beta \quad \text{plus-points } (A\text{-points}),$$
$$m + \alpha + \beta + k - \Delta \quad \text{minus-points } (B\text{-points});$$

here $F(x, y) > 0$.

From lemma 4 it follows that besides these points there exist

$$\frac{(n-1)^2 - k + 1}{2} - p - \alpha - \beta - \gamma \quad \text{plus-points } (B'\text{-points}),$$

$$\frac{(n-1)^2 + k - 1}{2} - m - \alpha - \beta - \gamma - k + \Delta \quad \text{minus-points } (A'\text{-points}),$$

where $F(x, y) < 0$; 2γ again denotes here the number of imaginary critical points of $F(x, y)$.

At

$$\frac{(n-1)^2 + k - 1}{2} + p - m - \gamma - k + \Delta$$

A- and A'-points we have $F/J > 0$ while at

$$\frac{(n-1)^2 - k + 1}{2} - p + m - \gamma + k - \Delta$$

B- and B'-points we have $F/J < 0$. Hence it is easy to see that both of the numbers

$$\frac{(n-1)^2 + k - 1}{2} + p - m - k + \Delta$$

and

$$\frac{(n-1)^2 - k + 1}{2} - p + m + k - \Delta$$

exceed

$$\frac{\left(\frac{n-5}{2} + 1\right)\left(\frac{n-5}{2} + 2\right)}{2} - 1 = \frac{(n-1)(n-3)}{8} - 1.$$

In fact supposing the contrary we can so chose the $\frac{1}{8}(n-1)(n-3)$ coefficients of the function $f(x, y)$ not vanishing identically that it will vanish at all A- and A'-points (resp. at all B- and B'-points) and besides that the real parts of all complex members $P(x_i, y_i)/J(x_i, y_i)$ in (4) corresponding to the imaginary critical points of $F(x, y)$ vanish also. Lemma 6 apprises us that such a choice is possible. Using lemma 5 we can prove (in the same way as we proved an analogous assertion above; see the proof of the First Fundamental Theorem) that this function $f(x, y)$ differs from 0 in at least one real critical point of

$F(x, y)$. Then the left hand member of (4) will be the sum of real quantities having the same sign, one of which at least is $\neq 0$. This sum cannot be 0. Hence

$$\left| m - p - \Delta + \frac{k+1}{2} \right| \leqq \frac{3n^2 - 4n + 1}{8}, \tag{8}$$

which completes the proof of our theorem.

3. Discussion of the Results Obtained. Some Examples

First of all we must prove that *the limits obtained for* $|p - m|$, *viz.*

$$|p - m| \leqq \frac{3n^2 - 6n}{8} + 1$$

for n even and

$$\left| m - p - \Delta + \frac{k+1}{2} \right| \leqq \frac{3n^2 - 4n + 1}{8}$$

for n odd are the best possible.

To this end we shall construct for every even n an algebraic curve of order n consisting of $\frac{1}{8}(3n^2 - 6n) + 1$ ovals lying exterior to each other and for every odd n a curve of order n consisting of $\frac{1}{8}(3n^2 - 4n + 1) - \frac{1}{2}(n - 1)$ positive ovals lying exterior to each other and all in the one and the same region bounded by the odd branch and the line at infinity, and besides such that the number of the regions G_i, defined in lemma 3, will be equal (for this curve) to n. The process of constructing these curves will be a slight modification of Harnack's construction of M-curves. We shall prove the existence of such curves inductively.

First of all for $n = 1$ and $n = 2$ these curves exist.

Suppose now that for a certain *even* n there exists a curve $a^{(n)} = 0$ of order n consisting of $\frac{1}{8}(3n^2 - 6n) + 1$ ovals lying exterior to each other. Suppose moreover that one of these ovals intersects a straight line $a^{(1)} = 0$ in two real points $A_n^{(1)}$ and $A_n^{(n)}$ and touches the same line at $\frac{1}{2}(n - 2)$ points $A_n^{(2i)} \equiv A_n^{(2i+1)}$ $(i = 1, 2, \cdots, \frac{1}{2}(n - 2))$; we suppose that the order of the points $A^{(1)}, A^{(2i)}, A^{(n)}$ on $a^{(1)} = 0$ coincides with the order of their indices. In the initial case $n = 2$ we can take for $a^{(2)} = 0$ an ellipse meeting a line $a^{(1)} = 0$ in two different real points.

Consider the curve

$$a^{(n+1)} \equiv a^{(1)}a^{(n)} + \lambda q^{(n+1)} = 0$$

where λ is a certain real constant and $q^{(n+1)} = 0$ is the equation (with real coefficients) of an algebraic curve of order $n + 1$ without real singularities

which meets the line $a^{(1)} = 0$ in $n + 1$ different real points $A^{(1)}_{n+1}, A^{(2)}_{n+1}, \cdots A^{(n+1)}_{n+1}$ all lying outside the segment $A^{(1)}_n A^{(2)}_n A^{(n)}_n$.[14]

Then taking λ sufficiently small we can choose its sign in such a way that the curve

$$a^{(n+1)} = 0$$

consists of

1) an odd branch meeting the line $a^{(1)} = 0$ in $n + 1$ points $A^{(1)}_{n+1}, A^{(2)}_{n+1}, \cdots, A^{(n+1)}_{n+1}$ and

2) of

$$\frac{3n^2 - 6n}{8} - \frac{n}{2} = \frac{3n^2 - 2n}{8} = \frac{3(n+1)^2 - 4(n+1) + 1}{8} - \frac{(n+1) - 1}{2}$$

ovals which all lie in the same region defined on the projective plane by the odd branch of the curve $a^{(n+1)} = 0$ and the line $a^{(1)} = 0$. Besides when all these

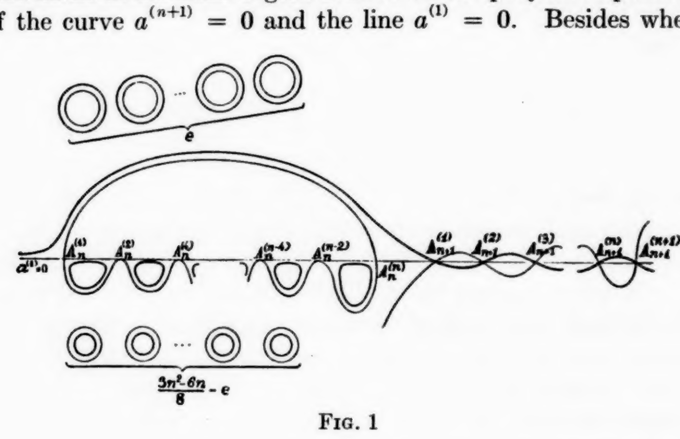


FIG. 1

ovals are positive, then $n + 1$ is the number of the regions G_i (defined as in lemma 3) bounded by the line $a^{(1)} = 0$ (which plays here the part of the line at infinity) and the odd branch of the curve $a^{(n+1)} = 0$. All these constructions are represented schematically in fig. 1, where the dark line represents the curve $a^{(n+1)} = 0$ and the light curves represent the curves $a^{(1)} = 0$, $a^{(n)} = 0$ and $q^{(n+1)} = 0$.[15]

---

[14] In order to construct such a curve take any real algebraic curve of order $n + 1$ meeting the line $a^{(1)} = 0$ in $n + 1$ real different points none of which lies on the segment $A^{(1)}_n A^{(2)}_n A^{(n)}_n$. If this curve has singularities we can vary its equation a little so that the new curve will be free of singularities; the displacement of the points of intersection of the curve with the line $a^{(1)} = 0$ will be arbitrarily small so that these points will remain real and distinct and will lie exterior to the segment $A^{(1)}_n A^{(2)}_n A^{(n)}_n$.

[15] We shall give here a detailed proof that for a sufficiently small λ different from 0 the curve $a^{(n+1)} = 0$ will have no singularities and will contain no ovals except those shown in fig. 1.

We begin with the first of these statements (viz. that the curve $a^{(n+1)} = 0$ will have no singular points). Let us call "critical" such values of λ that the curve $a^{(n+1)} \equiv a^{(1)} \cdot a^{(n)} +$

Suppose now that for a certain *odd* n there exists a curve $a^{(n)} = 0$ of order n consisting

1) of an odd branch intersecting the line $a^{(1)} = 0$ in n different real points

$$A_n^{(1)}, A_n^{(2)}, \cdots, A_n^{(n)},$$

2) of $\frac{1}{8}(3n^2 - 4n + 1) - \frac{1}{2}(n - 1)$ ovals all situated in the same region defined by the odd branch and the line $a^{(1)} = 0$ and lying exterior to each other.

Consider the curve

$$a^{(n+1)} \equiv a^{(1)} \cdot a^{(n)} + \lambda q_{n+1}^{(1)} \cdot q_{n+1}^{(2)} \cdots q_{n+1}^{(n+1)} = 0,$$

$\lambda q^{(n+1)} = 0$ has singular points. Evidently our statement will be proved if we prove that there is only a finite number of critical values of λ. A value of λ will be critical if it is a solution of the following equations (in x, y and λ)

$$(9) \qquad a^{(n+1)} = 0, \qquad \frac{\partial a^{(n+1)}}{\partial x} = 0 \quad \text{and} \quad \frac{\partial a^{(n+1)}}{\partial y} = 0.$$

Eliminating x and y, we obtain an algebraic equation for λ. Therefore in order to prove our statement (that there is only a finite number of critical values of λ) it is sufficient to demonstrate that these equations are not satisfied identically for all values of λ. This is evident since it is not satisfied for $\lambda = \infty$, the curve $q^{(n+1)} = 0$ having no singularities. Hence it follows that for λ_0 sufficiently small there are in the interval $(-\lambda_0, +\lambda_0)$ no critical values of λ with the exception of $\lambda = 0$; so that all curves $a^{(n+1)} = 0$, $0 < \lambda < \lambda_0$, have the same topological structure, and so do all curves $a^{(n+1)} = 0$, $-\lambda_0 < \lambda < 0$.

On the other hand it is evident that for small values of $|\lambda|$ the curve $a^{(n+1)} = 0$ passes "near" the curve $a^{(1)} \cdot a^{(n)} = 0$. Here the word "near" must be understood as meaning that if we project the projective plane on the hemisphere, the image of the curve $a^{(n+1)} = 0$ will lie in the neighbourhood (in the ordinary sense) of the image of the curve $a^{(1)} \cdot a^{(n)} = 0$. Evidently the curve $a^{(n+1)} = 0$ lies on that side of every portion of the curve $a^{(1)}a^{(n)} = 0$ for which $a^{(1)}a^{(n)}$ and $\lambda q^{(n+1)}$ have different signs. Take one of these portions and continue it until it intersects the curve $q^{(n+1)} = 0$. When we pass through a point $A_n^{(i)}$ we shall continue so as to be near the dark curve of fig. 1. In fig. 2 the segment AB of the curve $a^{(1)}a^{(n)} = 0$ is schematically represented; its endpoints A and B belong to the curve $q^{(n+1)} = 0$; the shaded part represents that part of the plane where $a^{(1)} \cdot a^{(n)}$ and $\lambda q^{(n+1)}$ have opposite signs. This region of the plane we shall denote by G. The black line represents the curve $a^{(n+1)} = 0$. If this curve (for a small $|\lambda|$) contains any other oval O_1 beside those represented in fig. 1 this oval must lie wholly within the region G (fig. 1) or another such region. Suppose now that λ approaches O. Two cases only are possible (because in the region G the expressions $a^{(1)} \cdot a^{(n)}$ and $q^{(n+1)}$ preserve their signs).

First case. The oval O_1 contracts. Then this oval cannot disappear without passing through a critical point. Therefore (since there are no critical values of λ within the interval considered) it must approach either an oval or a point which would have to be a singular point of the curve $a^{(1)}a^{(n)} = 0$. Both these situations are impossible.

Second case. The oval O_1 expands. Then it cannot approach the segment AB of the curve $a^{(1)} \cdot a^{(n)} = 0$ so as to merge with it when $\lambda = 0$.

We point out in conclusion that the ovals of the curve $a^{(n+1)} = 0$ which in fig. 1 lie within the corresponding ovals of the curve $a^{(n)} = 0$ can very well have the opposite mutual position i.e. the oval of the curve $a^{(n+1)} = 0$ can contain the oval of the curve $a^{(n)} = 0$ in its own interior. They can also intersect. This will happen if they are intersected by the curve $q^{(n+1)} = 0$ (they will intersect then in the points of intersection of the curve $q^{(n+1)} = 0$ with one of these ovals). The same remarks apply to all the other figures of this paper.

where $q_{n+1}^{(i)} = 0$ $(i = 1, \cdots, n + 1)$ are the equations of the straight lines intersecting the line $a^{(1)} = 0$ in the points $A_{n+1}^{(1)}, A_{n+1}^{(2)}, \cdots, A_{n+1}^{(n+1)}$ respectively which are all situated exterior to the segment $A_n^{(1)} A_n^{(2)} A_n^{(n)}$ and where λ is a constant. We shall also suppose that $q_{n+1}^{(2i)} \equiv q_{n+1}^{(2i+1)}$ for $i = 1, \cdots, \frac{1}{2}(n - 1)$, while all the points $A_{n+1}^{(1)}, A_{n+1}^{(2i)}, \cdots, A_{n+1}^{(n+1)}$ are different.

Then we can so chose the sign of λ that if we take λ sufficiently small numerically the curve $a^{(n+1)} = 0$ will consist of

$$\frac{3n^2 - 4n + 1}{8} - \frac{n - 1}{2} + n = \frac{3(n + 1)^2 - 6(n + 1)}{8} + 1$$

ovals all lying exterior to each other; the line $a^{(1)} = 0$ intersects one of these ovals in two points $A_{n+1}^{(1)}$ and $A_{n+1}^{(n+1)}$ and touches it in $\frac{1}{2}(n - 1)$ more points $A_{n+1}^{(2i)}$ $(i = 1, \cdots, \frac{1}{2}(n - 1))$, while it does not meet other ovals of the curve.

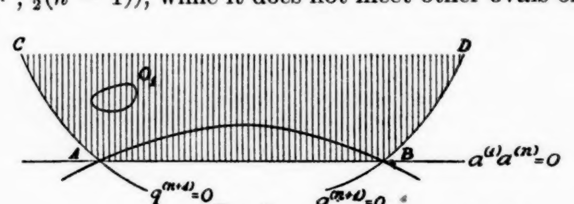


FIG. 2

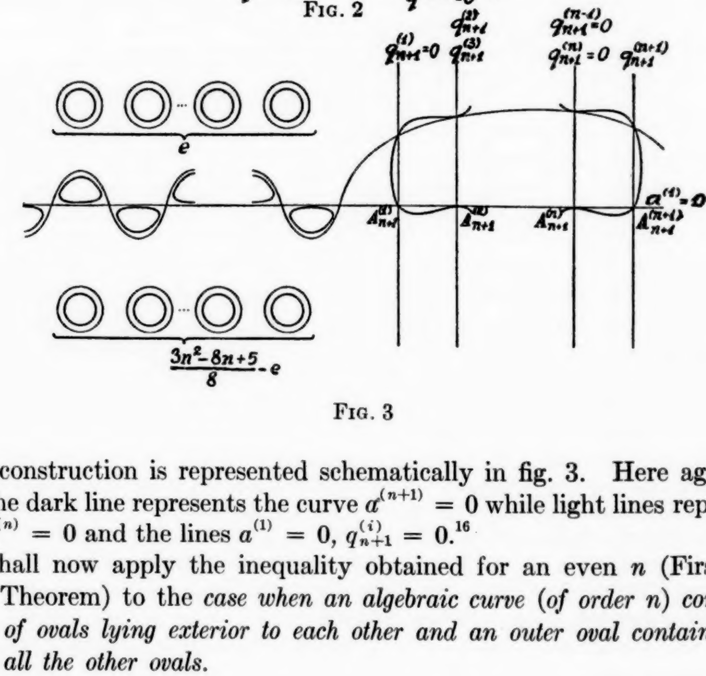


FIG. 3

This construction is represented schematically in fig. 3. Here again (as in fig. 1) the dark line represents the curve $a^{(n+1)} = 0$ while light lines represent the curve $a^{(n)} = 0$ and the lines $a^{(1)} = 0$, $q_{n+1}^{(i)} = 0$.[16]

We shall now apply the inequality obtained for an even n (First Fundamental Theorem) to the *case when an algebraic curve (of order n) consists of a number of ovals lying exterior to each other and an outer oval containing in its interior all the other ovals.*

[16] We can prove, just as we have done for the case of an even n, (see footnote 15) that if $|\lambda|$ is sufficiently small the curve $a^{(n+1)} = 0$ will have no singularities and will contain no ovals other than those shown by dark curves in fig. 3. Only here we see that the system (9) of equations has solutions for every λ using the theorem of Bertini.

Changing the sign of $F(x, y)$ if necessary we can make the outer oval positive; all other (inner) ovals will then be negative. Then in the First Fundamental Theorem we must make $\delta = 0$. Denoting by S the number of inner (negative) ovals we obtain (the number of positive ovals being 1) the inequality

$$S \leqq \frac{3n^2 - 6n}{8} + 1. \tag{10}$$

On the other hand for any $n = 4k + 2$ (where k is an integer $\geqq 0$) the algebraic curves of the order n exist consisting of an outer oval and of $S = \frac{1}{8}(3n^2 - 6n)$ ovals lying interior to it and exterior to each other.

In proving this we shall apply, with some modifications, Hilbert's process. The proof is inductive; suppose that for a certain $n = 4k + 2$ a curve $a^{(n)} = 0$ of order n exists consisting of an oval and in its interior $\frac{1}{8}(3n^2 - 6n)$ other

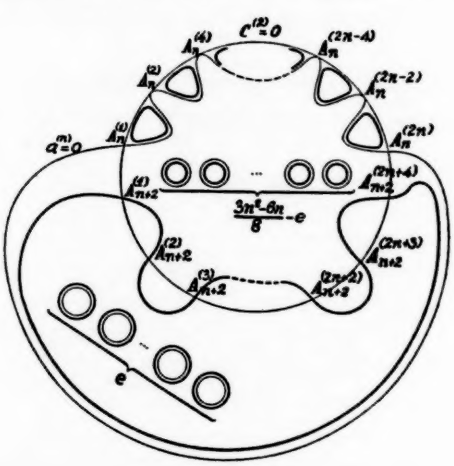


FIG. 4

ovals lying outside each other. We suppose besides that the outer oval intersects a certain ellipse $b^{(2)} = 0$ in two real points $A_n^{(1)}$ and $A_n^{(2n)}$ and touches it internally in $n - 1$ points $A_n^{(2)}, A_n^{(4)}, \cdots, A_n^{(2n-2)}$ coinciding, respectively with $A_n^{(3)}, A_n^{(5)}, \cdots, A_n^{(2n-1)}$ (see fig. 4, light line) and all lying on one of the two arcs of $b^{(2)} = 0$ having for its ends the points $A_n^{(1)}$ and $A_n^{(n)}$.

In the initial case $n = 2$ we must take for $a^{(2)} = 0$ an ellipse touching another ellipse $b^{(2)} = 0$ internally (at the point $A_n^{(2)}$) and meeting it besides in two other points $A_2^{(1)}$ and $A_2^{(4)}$.

On the arc $A_n^{(1)}A_n^{(2n)}$ of the ellipse $b^{(2)} = 0$ which does not contain the points $A_n^{(i)}$ take $2n + 4$ different points $A_{n+2}^{(1)}, \cdots, A_{n+2}^{(2n+4)}$ and draw through them a curve $q^{(n+2)} = 0$ (of order $n + 2$) without real singular points. Such curve can be obtained for instance if we vary a little the coefficients of the equation

$$q_{n+2}^{(1)} \cdot q_{n+2}^{(2)} \cdots q_{n+2}^{(n+2)} = 0,$$

where $q_{n+2}^{(i)} = 0$ (for $i = 1, \cdots, n + 2$) is the equation of a line which passes through the neighbouring points $A_{n+2}^{(i)}$. Then for a λ sufficiently small numerically and of a suitable sign the curve with the equation

$$a^{(n+2)} \equiv b^{(2)}a^{(n)} + \lambda q^{(n+2)} = 0$$

will resemble the curve represented by the dark line in fig. 4 and by the light line in fig. 5.

This curve consists of

1) $\dfrac{3n^2 - 6n}{8} - l + n$

ovals lying exterior to each other and within the ellipse $b^{(2)} = 0$. Here l denotes the number of ovals of the curve $a^{(n)} = 0$ lying outside the ellipse $b^{(2)} = 0$.

2) an oval meeting the ellipse $b^{(2)} = 0$ in $2n + 4$ different real points $A_{n+2}^{(i)}$; and

3) l ovals lying within the preceding oval and outside the ellipse.

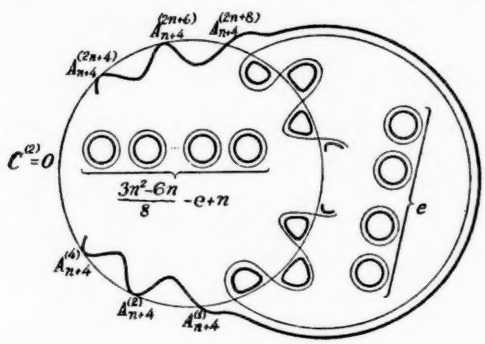


FIG. 5

That for $|\lambda|$ sufficiently small the curve $a^{(n+2)} = 0$ will contain no singular points and no other ovals than those represented in the figure can be proved in the same way as the analogous proposition was proved in footnote 15.

On the arc $A_{n+2}^{(1)}A_{n+2}^{(2n+4)}$ of the ellipse $b^{(2)} = 0$ containing no other points $A_{n+2}^{(i)}$ we shall take $2n + 8$ points

$$A_{n+4}^{(1)}, A_{n+4}^{(2)}, \cdots, A_{n+4}^{(2n+8)}$$

such that the points $A_{n+4}^{(2i)}$ and $A_{n+4}^{(2i+1)}$ $(i = 1, 2, \cdots, n + 3)$ coincide while all the points $A_{n+4}^{(1)}, A_{n+4}^{(2)}, A_{n+4}^{(4)}, \cdots, A_{n+4}^{(2n+6)}, A_{n+4}^{(2n+8)}$ are different (fig. 5). Now join the points $A_{n+4}^{(1)}$ and $A_{n+4}^{(2)}$; $A_{n+4}^{(3)}$ and $A_{n+4}^{(4)}$; $\cdots$; $A_{n+4}^{(2n+7)}$ and $A_{n+4}^{(2n+8)}$ by the lines

$$q_{n+4}^{(1)} = 0; \qquad q_{n+4}^{(2)} = 0; \cdots ; \qquad q_{n+4}^{(n+4)} = 0.$$

Then if λ is sufficiently small numerically and its sign is conveniently chosen, the curve

$$a^{n+4} \equiv b^{(2)}a^{(n+2)} + \lambda q_{n+4}^{(1)} \cdot q_{n+4}^{(2)} \cdots q_{n+4}^{(n+4)} = 0$$

(fig. 5, dark curve) consists of

$$\frac{3n^2 - 6n}{8} + n + 2n + 3 = \frac{3(n + 4)^2 - 6(n + 4)}{8}$$

ovals all lying exterior to each other and interior to another oval which intersects the ellipse $b^{(2)} = 0$ in two points $A_{n+4}^{(1)}$ and $A_{n+4}^{(2n+8)}$ and touches it internally at $(n + 3)$ other points $A_{n+4}^{(2i)}$ $(i = 1, \cdots, n + 3)$ which all lie on the same arc $A_{n+4}^{(1)} A_{n+4}^{(2n+8)}$ of the ellipse. Consequently our curve $a^{(n+4)} = 0$ of order $n + 4$ possesses all the required properties, q.e.d.[17]

K. Rohn proves in his work already mentioned which deals with the curves of order 6 that such a curve cannot consist of an outer oval and ten other ovals lying interior to it and exterior to each other. This proves that the upper bound of the number of such ovals which we have found viz.

$$\frac{3n^2 - 6n}{8} + 1$$

cannot be attained in the case $n = 6$, as well as in the case $n = 2$.

When $n = 4k$ (k an integer) there exist algebraic curves of order n consisting of an outer oval and

$$S = \frac{3n^2 - 6n}{8} - 2$$

other ovals lying exterior to each other but within the outer oval.

To start the induction we begin with the case $k = 1$ $(n = 4)$. Let $a^{(2)} = 0$ and $b^{(2)} = 0$ be the equations of two ellipses, of which the second lies in the interior of the first. Take 5 different points $A_4^{(1)}$, $A_4^{(2)} \equiv A_4^{(3)}$, $A_4^{(4)} \equiv A_4^{(5)}$, $A_4^{(6)} \equiv A_4^{(7)}$ and $A_4^{(8)}$ on the ellipse $b^{(2)} = 0$ and draw four lines through the points $A_4^{(1)}$ and $A_4^{(2)}$ $(q_4^{(1)} = 0)$, $A_4^{(3)}$ and $A_4^{(4)}$ $(q_4^{(2)} = 0)$; $A_4^{(5)}$ and $A_4^{(6)}$ $(q_4^{(3)} = 0)$ and $A_4^{(7)}$ and $A_4^{(8)}$ $(q_4^{(4)} = 0)$. Then the curve

$$a^{(4)} \equiv a^{(2)} b^{(2)} + \lambda q_1 q_2 q_3 q_4 = 0$$

(if λ is sufficiently small numerically and has the suitable sign) consists of two ovals (one of which lies within the other) differing but little from the two ellipses $a^{(2)} = 0$ and $b^{(2)} = 0$. The inner oval intersects the ellipse $b^{(2)} = 0$ in two points $A_4^{(1)}$ and $A_4^{(8)}$ and touches it from without at three points $A_4^{(2)}$, $A_4^{(4)}$, $A_4^{(6)}$ which all lie on the same arc $A_4^{(1)} A_4^{(8)}$ of the ellipse.

Suppose now that for a certain $n = 4k$ we have constructed an algebraic curve $a^{(n)} = 0$ of order n consisting of an outer oval and of $\frac{1}{8}(3n^2 - 6n) - 2$ other ovals lying within the outer oval and outside each other; one of these ovals intersects an ellipse $b^{(2)} = 0$ in two points $A_n^{(1)}$ and $A_n^{(2n)}$ and touches it from without in $n - 1$ points $A_n^{(2)}, A_n^{(4)}, \cdots, A_n^{(2n-2)}$ (coinciding respectively

[17] That for $|\lambda|$ sufficiently small the curve $a^{(n+4)} = 0$ contains no other ovals besides those indicated in fig. 5 and has no singularities, can be proved in the same way as the analogous statements in footnotes 15 and 16.

with $A_n^{(3)}, A_n^{(5)}, \cdots, A_n^{(2n-1)}$) all lying on the same arc $A_n^{(1)}A_n^{(n)}$ of the ellipse (fig. 6, light curve).

Take, on the other arc $A_n^{(1)}A_n^{(2n)}$, (i.e. on the arc $A_n^{(1)}A_n^{(2n)}$ which does not contain $A_n^{(2)}, \cdots, A_n^{(2n-2)}$) $2n + 4$ different points $A_{n+2}^{(1)}, A_{n+2}^{(2)}, \cdots, A_{n+2}^{(2n+4)}$. Through these points we draw an algebraic curve of $q^{(n+2)} = 0$ order $n + 2$ with-

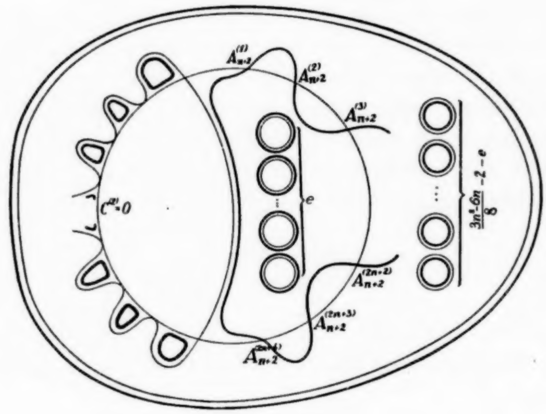


FIG. 6

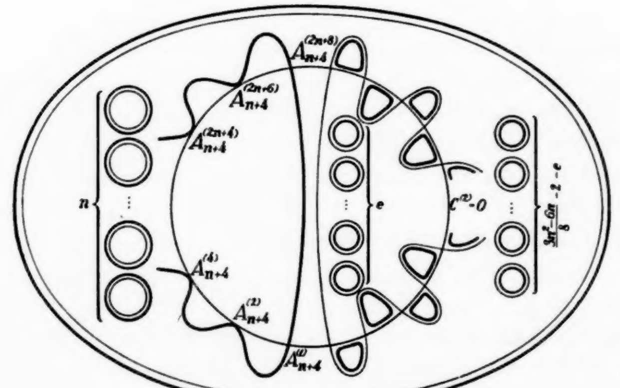


FIG. 7

out singularities. Then (if $|\lambda|$ is taken sufficiently small and the sign of λ is suitably chosen) the curve

$$a^{(n+2)} \equiv a^{(n)} \cdot b^{(2)} + \lambda q^{(n+2)} = 0$$

will have the form like that indicated schematically in fig. 6 (dark curve) and fig. 7 (light curve).

Now take on the arc $A_{n+2}^{(1)}A_{n+2}^{(2n+4)}$ of the ellipse $b^{(2)} = 0$ which does not contain any other points $A_{n+2}^{(i)}$, $n + 5$ different points $A_{n+4}^{(1)}$, $A_{n+4}^{(2)} \equiv A_{n+4}^{(3)}$, $A_{n+4}^{(4)} \equiv A_{n+4}^{(5)}, \cdots, A_{n+4}^{(2n+6)} \equiv A_{n+4}^{(2n+7)}$, $A_{n+4}^{(2n+8)}$ and draw the lines $q_{n+4}^{(1)} = 0, q_{n+4}^{(2)} = 0, \cdots, q_{n+4}^{(n+4)}$ through $A_{n+4}^{(1)}$ and $A_{n+4}^{(2)}$, $A_{n+4}^{(3)}$ and $A_{n+4}^{(4)}, \cdots,$

$A_{n+4}^{(2n+7)}$ and $A_{n+4}^{(2n+8)}$, resp. Then (if λ is sufficiently small numerically and of a suitable sign) the curve of order $n + 4$ $a^{(n+4)} \equiv a^{(n+2)}b^{(2)} + \lambda q_{n+4}^{(1)} \cdot q_{n+4}^{(2)} \cdots q_{n+4}^{(n+4)}$ (see fig. 7, dark curve) consists of an outer oval and

$$\frac{3n^2 - 6n}{8} - 2 + n + 2n + 4 - 1 = \frac{3(n + 4)^2 - 6(n + 4)}{8} - 2$$

inner ovals (all lying within the outer oval and outside each other) one of which intersects the ellipse $b^{(2)} = 0$ in two points $A_{n+4}^{(1)}$ and $A_{n+4}^{(2n+8)}$ and touches it at $n + 3$ other points $A_{n+4}^{(2)}, A_{n+4}^{(4)}, \cdots, A_{n+4}^{(2n+6)}$ (coinciding resp. with $A_{n+4}^{(3)}, A_{n+4}^{(5)}, \cdots, A_{n+4}^{(2n+7)}$) which all lie in the same arc $A_{n+4}^{(1)}A_{n+4}^{(2n+8)}$.[18]

The analogous constructions as well as the inequalities analogous to (10) can of course be effected for the case of an odd n.

I shall conclude this paper with the remark that in all examples that I know not only do we have

$$|p - m| \leqq \frac{3n^2 - 6n}{8} + 1$$

for even n and

$$|p - m| \leqq \frac{3n^2 - 4n + 1}{8} + \left|\frac{k + 1}{2} - \Delta\right|$$

for odd n, but the same inequalities are true for p and m separately. We could prove this to be so if there existed a method which allowed us to destroy an arbitrary oval of a curve of order n without affecting either its order or the topology of other ovals. All curves constructed by Harnack's or Hilbert's process admit of such a wiping-out of their ovals. In fact the examples given above are exactly such Harnack or Hilbert curves with some of their ovals wiped-out. But whether it is possible to destroy ovals (without affecting the topology of other ovals) of every algebraic curve I do not know.

In the same manner we could prove that S does not exceed $\frac{1}{8}(3n^2 - 6n)$ if we could make the outer oval (expanding) touch itself without affecting the topology of other ovals. Then either a new negative oval would appear or the outer positive oval would become negative. In both cases the inequality $S \leqq \frac{1}{8}(3n^2 - 6n)$ would be established.

I wish to thank E. M. Livenson for the help he has given me in the editing of this paper.

Moscow, U. S. S. R.

---

[18] In the same manner as was done in the case of $n = 4k + 2$ we can prove that the curve $a^{(n+4)} = 0$ will not have a more complicated topological structure (i.e., it will not have any ovals other than those represented in fig. 7, and it will not have singular points).

ANNALS OF MATHEMATICS
Vol. 39, No. 1, January, 1938

ARITHMETICAL PROPERTIES OF SEQUENCES IN RINGS

BY MORGAN WARD

(Received February 1, 1937; Revised June 21, 1937)

INTRODUCTION

1. Let S be the set of numbers $0, 1, 2, \cdots, \mathfrak{O}$, a commutative ring of elements $A, B, \cdots, Z, \cdots$ containing a unit element, and let U_n be a one-valued function on S to $\mathfrak{O}$; that is, a sequence

$$(U): \qquad U_0, U_1, \cdots, U_n, \cdots$$

of elements of $\mathfrak{O}$. The object of this paper is to study the periodicity and divisibility of such sequences relative to ideals of $\mathfrak{O}$. If we extend[1] U_n over the ring of rational integers by letting $U_{-n} = U_n$, we have a special instance of a correspondence between a commutative ring and a structure (lattice) studied in a previous paper in these Annals (Ward [1][2]).

The less general hypotheses of the present paper allow us to prove as theorems many of the axioms assumed in W. A. The results of the paper are however of a quite different character from those in W. A., and the paper may be read independently.

In conjunction with W. A. this paper gives a general theory of the arithmetic properties of sequences which renders any half-hearted generalizations of the ordinary theory of linear sequences of rational integers[3] such as to linear sequences of algebraic integers, to a large extent superfluous. In addition, we obtain many results for the special case of linear sequences of rational integers under much less restrictive hypotheses than heretofore.[4]

The existence of a smaller period for the places of apparition of a divisor of a linear divisibility sequence than the restricted period of the sequence which we prove here abstractly[5] is a fact of some arithmetical interest which does not seem to have been observed previously.

---

[1] We include the case when U_n is defined over a sub-set T of S by letting U_n be the zero of $\mathfrak{O}$ over the complementary set $S - T$. For example a function defined merely for $n = 0$, 1 and 2 is regarded as a sequence in which all terms vanish after the third.

[2] The bracketed numerals refer to the reference listed at the close of the paper. I shall refer to this particular paper as W. A.

[3] See W. A. for references to recent papers.

[4] See for example theorems 3.2, 4.1, 4.2 and 4.3. Theorem 3.2 is given in Ward [2] for a very special case. See also Carmichael [1]. Theorems 4.1–4.3 are closely connected with Marshall Hall's Theorem III (Hall [1], p. 579).

[5] Numerical examples over the ring of rational integers can be constructed without much difficulty, but unfortunately satisfy difference equations of quite high order.

2. We shall adhere to the following scheme of notation based on van der Waerden [1] and used in W. A. Elements of $\mathfrak{O}$ are denoted by roman capitals; small italic and Greek letters denote ordinary integers. The ideals of $\mathfrak{O}$ are denoted by German letters $\mathfrak{A}, \mathfrak{B}, \cdots$ $(\mathfrak{A}, \mathfrak{B})$ and $[\mathfrak{A}, \mathfrak{B}]$ denote the union and cross-cut of the ideals $\mathfrak{A}$ and $\mathfrak{B}$; (a, b) and $[a, b]$ the greatest common divisor and least common multiple of the numbers a and b; (A, B) the union of the principal ideals $[A]$ and $[B]$. The letters U and V are reserved to denote sequences. Thus (U) stands for a function. If the ideals $\mathfrak{A}$ and $[A]$ are co-prime ("teilerfremd," van der Waerden [1], §85), we write $(\mathfrak{A}, A) = \mathfrak{O}$. a divides b is written as usual $a \mid b$.

II. Modular Periodicity and Divisibility Sequences

3. We begin with some definitions. The sequence (U) is said to be *finite* if it contains only a finite number of non-vanishing terms. It is said to be *linear over* $\mathfrak{O}$ if its terms satisfy a recursion relation

$$U_{n+k} = C_1 U_{n+k-1} + \cdots + C_k U_n, \qquad (n = 0, 1, 2, \cdots) \tag{3.1}$$

with coefficients C_i in $\mathfrak{O}$.

(U) is said to be a *divisibility sequence* (M. Hall [1], Ward [1], [4]) if U_n divides U_m in $\mathfrak{O}$ whenever n divides m. If (U) is both a linear sequence and a divisibility sequence we shall call (U) "Lucasian"[6] in honor of the French mathematician E. Lucas who first systematically studied a special class of such sequences. (Lucas [1], [2], Dickson [1].)

It may happen that the terms of (U) become periodic when taken to a fixed ideal modulus $\mathfrak{A}$ of $\mathfrak{O}$; that is, there exist numbers λ and ν such that

$$U_{n+\lambda} \equiv U_n \qquad (\text{mod } \mathfrak{A}),\ n \geqq \nu. \tag{3.2}$$

The least such λ and ν are called the *period* and *numeric* of (U) for the modulus $\mathfrak{A}$. This minimal period is easily seen to divide every other period. If $\nu = 0$, (U) is said to be *purely periodic* modulo $\mathfrak{A}$.

If there exists at least one term U_k of (U) such that $U_k \equiv 0 \pmod{\mathfrak{A}}$ then $\mathfrak{A}$ is called a divisor of (U). If all terms of (U) are divisible by $\mathfrak{A} \neq \mathfrak{O}$ from a certain point on, then $\mathfrak{A}$ is called a *null divisor* of (U), and (U) a null sequence modulo $\mathfrak{A}$. Every finite sequence is thus a null sequence for any modulus.

A positive integer μ is said to be a *restricted period* of (U) modulo $\mathfrak{A}$ if there exists an element A of $\mathfrak{O}$ such that

$$U_{n+\mu} \equiv A U_n \qquad (\text{mod } \mathfrak{A}), \text{ all large } n.^7 \tag{3.3}$$

Here A depends on μ. We call A a multiplier of (U) modulo $\mathfrak{A}$. The least μ for which (3.3) holds is called *the* restricted period of (U).

---

[6] The more euphonious term "Lucas sequence" already has a precise meaning in the literature.

[7] It usually suffices to consider (3.3) for n greater than the numeric of (U) modulo $\mathfrak{A}$.

THEOREM 3.1. *Let (U) be a sequence of $\mathfrak{O}$, and $\mathfrak{A}$ any ideal of $\mathfrak{O}$ such that no divisor of $\mathfrak{A}$ is a null divisor of (U). Then if (U) is periodic modulo $\mathfrak{A}$, the minimal restricted period μ of (U) modulo $\mathfrak{A}$ exists, and divides every other restricted period, and in particular the actual period λ. Furthermore the multipliers of (U) modulo $\mathfrak{A}$ are all prime to $\mathfrak{A}$,[8] and form a group with respect to multiplication modulo $\mathfrak{A}$.*

PROOF. If $\mathfrak{A}$ is a null divisor of (U), (3.3) becomes a triviality. In any event, if (U) is periodic modulo $\mathfrak{A}$, the actual period λ is a restricted period with $A = I$. Hence a minimal μ exists $\leqq \lambda$, and we may write $\lambda = s\mu - t$ where $s \geqq 1$, $0 \leqq t < \mu$. Then for all large n,

$$U_{n+t} \equiv U_{n+t+\lambda} \equiv U_{n+s\mu} \equiv A^s U_n \pmod{\mathfrak{A}}$$

Hence $t = 0$ by the minimal property of μ, and

$$(A^s - 1)U_n \equiv 0 \quad (\mathrm{mod}\ \mathfrak{A}), \text{ all large } n. \tag{3.4}$$

I say that

$$(A, \mathfrak{A}) = \mathfrak{O}. \tag{3.41}$$

For if $(A, \mathfrak{A}) = \mathfrak{B} \neq \mathfrak{O}$, then $(A^s - 1, \mathfrak{B}) = \mathfrak{O}$ so that by (3.4), $U_n \equiv 0 \pmod{\mathfrak{B}}$, all large n, contradicting the hypothesis that no divisor of $\mathfrak{A}$ is a null divisor of (U).

Let ϕ be any other restricted period with multiplier B, and write $\phi = u\mu + \theta$, $u \geqq 1$, $0 \leqq \theta < \mu$. Then by (3.3)

$$A^u U_{n+\theta} \equiv U_{n+\theta+u\mu} \equiv U_{n+\phi} \equiv BU_n \quad (\mathrm{mod}\ \mathfrak{A}). \tag{3.5}$$

Therefore

$$(B, \mathfrak{A}) = \mathfrak{O}. \tag{3.51}$$

For if $(B, \mathfrak{A}) = \mathfrak{B} \neq \mathfrak{O}$, then by (3.5) and (3.41) $U_{n+\theta} \equiv 0 \bmod \mathfrak{B}$ for all large n, contradicting the hypothesis that no divisor of $\mathfrak{A}$ is a null divisor of (U).

Now the set of all elements of $\mathfrak{O}$ which are prime to $\mathfrak{A}$ form a group with respect to multiplication modulo $\mathfrak{A}$. (van der Waerden [1], Chapter XII.) Hence by (3.41) there exists an element A' of $\mathfrak{O}$ such that $A'A \equiv I \pmod{\mathfrak{A}}$ Thus by (3.5)

$$U_{n+\theta} \equiv (A'A)^u U_{n+\theta} \equiv A'^u BU_n \quad (\mathrm{mod}\ \mathfrak{A}).$$

Therefore $\theta = 0$ by the minimal property of μ, and μ divides ϕ.

It remains to prove the group property of the multipliers. It follows from (3.51) that the multipliers of (U) form a semi-group. All that remains is to show the existence of an inverse for each multiplier B.

---

[8] This statement is taken as an axiom in the discussion of the restricted period in part IV of W. A.

Since $(B, \mathfrak{A}) = \mathfrak{O}$, there exists an element B' of $\mathfrak{O}$ such that $BB' \equiv 1 \pmod{\mathfrak{A}}$ while $U_{n+\phi} \equiv BU_n \pmod{\mathfrak{A}}$, $n \geqq \nu$. On replacing n by $n - \phi$, we obtain for $n \geqq \nu + \phi$

$$U_{n-\phi} \equiv (B'B)U_{n-\phi} \equiv B'U_n \qquad \pmod{\mathfrak{A}}.$$

Determine positive integers x, y such that $x\phi = y\lambda$ where λ as usual is the period of (U). Then

$$U_{n+(x-1)\phi} \equiv U_{n-\phi+y\lambda} \equiv U_{n-\phi} \equiv B'U_n \qquad \pmod{\mathfrak{A}}.$$

Hence B' is a multiplier. This proof fails if $x = 1$. But then $\phi = \lambda$, $B \equiv I \pmod{\mathfrak{A}}$ so that $B' \equiv I \pmod{A}$ directly.

THEOREM 3.2. *Let $\mathfrak{O}$ be a ring in which the chain condition ("Teilerkettenforderung") holds for ideals. Let (U) be a sequence of $\mathfrak{O}$ and $\mathfrak{A}$ an ideal such that (U) is periodic modulo $\mathfrak{A}$, but such that no divisor of $\mathfrak{A}$ is a null divisor of (U). Then if λ is the period and μ the restricted period of (U) modulo $\mathfrak{A}$, the multipliers of (U) form a cyclic group of order λ/μ. Furthermore the multiplier A of* (3.3) *associated with the restricted period is a generator of this group.* (Ward [2].)

PROOF. Consider the sequence of ideals

$$\mathfrak{A}_0 = (\mathfrak{A}, U_r), \qquad \mathfrak{A}_1 = (\mathfrak{A}, U_r, U_{r+1}), \qquad \mathfrak{A}_2 = (\mathfrak{A}, U_r, U_{r+1}, U_{r+2}), \cdots$$

where r is a fixed number greater than the numeric of (U). Then $\mathfrak{A}_{i+1} \supset \mathfrak{A}_i$, $(i = 0, 1, 2, \cdots)$. Therefore by the chain condition, all the $\mathfrak{A}_i$ are equal from a certain point on. This resulting ideal $\mathfrak{T}$ divides both $\mathfrak{A}$ and every term of (U) beyond a certain point. Since (U) has no null divisors dividing $\mathfrak{A}$, $\mathfrak{T} = \mathfrak{O}$. Thus for some number l,

$$(\mathfrak{A}, U_r, U_{r+1}, \cdots U_{r+l-1}) = \mathfrak{O}.$$

It follows that the ideal $(U_{r+1}, U_{r+1}, \cdots U_{r+l-1})$ contains a number

$$W = X_1U_r + \cdots + X_lU_{r+l-1}, \qquad X \text{ in } \mathfrak{O} \tag{3.6}$$

such that $(W, \mathfrak{A}) = \mathfrak{O}$.

With the notation of the previous theorem, choose r so that the congruences (3.4) and (3.5) hold for $n \geqq r$. Then by (3.6) $(A^s - 1)W \equiv (B - A^u)W \equiv 0 \pmod{\mathfrak{A}}$, or $A^s \equiv 1$, $B \equiv A^u \pmod{\mathfrak{A}}$. Now if we define s as the *least* integer such that $A^s \equiv 1 \pmod{\mathfrak{A}}$ the multipliers are seen to form a cyclic group of order s with A as a generator. From the minimal property of λ, $\lambda = \mu s$ and $s = \lambda/\mu$.

4. The following easily proved theorem on the divisors of any sequence extends a previous theorem of mine (Ward [1] theorem 5.3) and is the basis for the study of divisors of divisibility sequences.

THEOREM 4.0. *Let (U) be a sequence over $\mathfrak{O}$, and $\mathfrak{M}$ any divisor of (U). Then if $\mathfrak{M} = [\mathfrak{A}, \mathfrak{B}]$, the set of places of apparition of $\mathfrak{M}$ in (U) is the cross-cut of the sets of places of apparition of $\mathfrak{A}$ and $\mathfrak{B}$ in (U).*

If in particular (U) is a divisibility sequence, the places of apparition of any divisor of (U) have the property of being closed under multiplication by positive integers. Furthermore, for any place of apparition s of a divisor $\mathfrak{A}$ of (U) there will exist a number r dividing s and such that

$$(4.1) \qquad U_r \equiv 0 \pmod{\mathfrak{A}}, \qquad U_x \not\equiv 0 \pmod{\mathfrak{A}} \text{ if } x \mid r.$$

Following M. Hall [1], we call such a number r a *rank of apparition*[9] of $\mathfrak{A}$ in (U). Of paramount interest and simplicity are the cases when a divisor of (U) has only a *finite* number of ranks of apparition. We then say the ranks of apparition constitute a *multiplicative set*. The theory of such sets is developed in part III of this paper. In the present section, we give a series of theorems on the finiteness of the ranks of apparition.

THEOREM 4.1. *If $\mathfrak{A}$ is a divisor of the divisibility sequence (U) and if (U) is also periodic modulo $\mathfrak{A}$, then a necessary and sufficient condition that $\mathfrak{A}$ shall have only a finite number of ranks of apparition are that all its ranks of apparition divide the restricted period of (U) modulo $\mathfrak{A}$.*

PROOF. The sufficiency of this condition is obvious. To establish its necessity, let r be a rank of apparition of $\mathfrak{A}$ which does not divide the restricted period μ, and let $(r, \mu) = d \not\equiv r$. For any positive integer t, we can choose positive integers x_t, y_t such that $td = y_t r - x_t \mu$. Then if $t \geqq \nu/d$ (where ν is the numeric of (U)),

$$U_{x_t \mu + td} \equiv A^{x_t} Y_{td} \equiv U_{y_t r} \equiv 0 \qquad \pmod{\mathfrak{A}},$$

for (U) is a divisibility sequence and $U_r \equiv 0 \pmod{\mathfrak{A}}$. By theorem 3.2, $(A, \mathfrak{A}) = \mathfrak{O}$, so that $U_{td} \equiv 0 \pmod{\mathfrak{A}}$. Hence td is a place of apparition of $\mathfrak{A}$ in (U) and is hence divisible by one or more ranks of apparition r' of $\mathfrak{A}$. If t is a prime number, r' must be divisible by t. For otherwise $r' \mid td$ implies $r' \mid d$, so that $r' \mid r$, $r' = r$, $d = r$. Hence

$$(4.2) \qquad td \geqq r' \geqq t \qquad t \text{ a prime.}$$

Since the number of primes is infinite, we may choose an infinite sequence of primes $t_0, t_1, t_2, \cdots$ such that $t_{n+1} > t_n d$, $t_0 d \geqq \nu$. Then the inequality (4.2) implies the existence of an infinity of ranks of apparition.

THEOREM 4.2. *Let (U) be a divisibility sequence, and $\mathfrak{A}$ a divisor of (U) such that (U) is purely periodic modulo $\mathfrak{A}$. Then $\mathfrak{A}$ has only a finite number of ranks of apparition and each such rank divides the restricted period of (U) modulo $\mathfrak{A}$.*

PROOF. Let r be a rank of apparition of $\mathfrak{A}$ in (U). In view of the previous theorems we need only show that r divides μ. Let $(r, \mu) = d$. Then it suffices to show that d is a place of apparition of $\mathfrak{A}$, for then since $d \mid r$, $r = d$ and $r \mid \mu$.

---

[9] In Ward [1], [3], [4], a rank of apparition was defined by the stricter requirement $U_r \equiv 0 \pmod{\mathfrak{A}}$, $U_x \not\equiv 0 \pmod{\mathfrak{A}}$ if $0 < x < r$, or in the case considered in Ward [1], the places of apparition were required to form an ideal. Although such a definition leads to many interesting results and is apparently met with frequently in the numerical cases of the Lucasian sequences from which the theory springs, (4.1) appears preferable.

Now there exist positive integers x, y such that $d = rx - \mu y$. Then by (3.3) and our hypotheses on r and (U),

$$A^{\nu} U_d \equiv U_{d+\mu\nu} \equiv U_{rx} \equiv 0 \qquad (\text{mod } \mathfrak{A}).$$

But $(A, \mathfrak{A}) = \mathfrak{O}$. Hence $U_d \equiv 0 \ (\text{mod } \mathfrak{A})$.

III. Multiplicative Sets

5. Let $r_1, r_2, \cdots, r_n$ be n fixed numbers. The set M consisting of all their integral multiples will be called the *multiplicative set based on* $r_1, r_2, \cdots, r_n$. If r_i divides r_j only for $i = j$, $(i, j = 1, \cdots, n)$, the r_i will be called the *generators* of the set. A generator is thus any element of the set which is irreducible in the set. Henceforth we assume that $r_1, \cdots, r_n$ are a set of generators of M.

The multiplicative set based on $r_1, \cdots, r_n$ is thus the maximal multiplicative semi-group containing $r_1, \cdots, r_n$ as its only irreducible elements.

The number $r = [r_1, r_2, \cdots, r_n]$ (where here and later $[x, \cdots, z]$ denotes the least common multiple or L.C.M. of the numbers $x, \cdots, z$) is called the *rank* of the set M.

Theorem 5.1. *If r is the rank of the multiplicative set M, every element of M is congruent modulo r to an element of the set greater than or equal to zero and less than r.*

Proof. If x lies in M, there exists a generator r_i dividing x: $x = yr_i$. Also $r = zr_i$ by the definition of L.C.M. Hence if $x = qr + t$ where $0 \leqq t < r$, then $t \equiv 0 \ (\text{mod } r_i)$ so that t lies in M and $x \equiv t \ (\text{mod } r)$.

We call a set of distinct elements of M which lie in a complete residue system modulo r a *representative set* of M.

Theorem 5.2. *The number of elements in a representative set of M is given by the formula*

$$r \sum_{s=1}^{n} (-1)^{s-1} \sum_{(i)} \frac{1}{[r_{i_1}, r_{i_2}, \cdots r_{i_s}]}.$$

Here the inner summation is taken over all the $\binom{n}{s}$ distinct combinations $i_1, \cdots, i_s$ of the subscripts 1 to n of the generators r_i taken s at a time. We omit the (simple) proof of this theorem here.[10]

Theorem 5.3. *The cross-cut of any two multiplicative sets is a multiplicative set, and each generator of the cross-cut is the* L.C.M. *of generators of the component sets.*

Proof. Let the sets be M_1 and M_2 and let $M_3 = [M_1, M_2]$ be their cross-cut. M_3 is obviously a multiplicative set. Every element of M_3 lies both in M_1

[10] For example take $r_1 = 6$, $r_2 = 10$, $r_3 = 15$. Then $r = 30$. A representative set consists of the eight numbers 0, 6, 10, 12, 15, 18, 20 and 24. The formula gives

$$30\left\{\left(\frac{1}{6} + \frac{1}{10} + \frac{1}{15}\right) - \left(\frac{1}{30} + \frac{1}{30} + \frac{1}{30}\right) + \left(\frac{1}{30}\right)\right\} = 5 + 3 + 2 - 3 + 1 = 8.$$

and M_2 and is hence divisible by a generator r_i of M_1 and a generator r'_j of M_2 and hence by their L.C.M. $[r_i, r'_j]$. Therefore M_3 is based on the set of nm elements $[r_1, r'_1], \cdots, [r_n, r'_m]$ where the r_i and r'_j are respectively the generators of M_1 and M_2. Hence the generators of M_3 consist of the irreducible elements in the set $[r_i, r'_j]$.

In like manner it is easy to prove

THEOREM 5.4. *The union of two multiplicative sets is a multiplicative set. If $r_1, \cdots, r_n$; $r'_1, \cdots, r'_m$ are the generators of the two sets, the generators of their union consist of the irreducible elements in the set of $n + m$ elements $r_1, \cdots, r'_m$.*

THEOREM 5.5. *The aggregate of all multiplicative sets forms an arithmetic structure* (O. Ore [1])——or (distributive) C-lattice—(G. Birkoff [1]) *with respect to the operations of forming the union and cross-cut.*[11]

PROOF. Let M_1, M_2, M_3 be any three multiplicative sets with generators $r_1, \cdots, r_n$; $r'_1, \cdots, r'_m$; $r''_1, \cdots, r''_p$. Let $[M_i, M_j]$, (M_i, M_j) stand for cross-cut and union respectively. Then it suffices to show that

$$[M_1, (M_2, M_3)] = ([M_1, M_2], [M_1, M_3]).$$

But this equality is obvious, since by the preceding two theorems, both sets are based upon the numbers $[r_1, r'_1], \cdots, [r_n, r'_m], [r_1, r''_1], \cdots [r_n, r''_p]$.

It is evident that the theorems of this section will also be true with slight changes of wording for multiplicative sets defined over a principal ideal ring $\mathfrak{O}$ all of whose quotient rings $\mathfrak{O}/\mathfrak{A}$ are finite.

IV. LINEAR SEQUENCES

6. THEOREM 6.1. *Let (U) be a linear sequence, and let $\mathfrak{A}$ be an ideal of $\mathfrak{O}$ whose quotient ring $\mathfrak{O}/\mathfrak{A}$ is of finite order. Then (U) is periodic modulo $\mathfrak{A}$. Furthermore, if $\mathfrak{A}$ is relatively prime to the last term C_k in the recursion* (3.1) *defining (U), then (U) is purely periodic modulo $\mathfrak{A}$.*

PROOF. (Ward [2].) The sequence (U) will become periodic modulo $\mathfrak{A}$ if any set of k residues of k consecutive terms of (U) modulo $\mathfrak{A}$ occurs more than once. Now the first $n + k - 1$ terms of (U) contain the $n + 1$ sets of k consecutive terms $U_0, \cdots, U_{k-1}; \cdots; U_n, \cdots U_{n+k-1}$. Let T be the order of the finite ring $\mathfrak{O}/\mathfrak{A}$, so that a complete residue system modulo $\mathfrak{A}$ contains T distinct elements. Then the period λ of (U) modulo $\mathfrak{A}$ is at most $T^k - 1$. Thus (3.2) holds with $\lambda \leqq T^k - 1$, $\nu \leqq T^k - 1$.

---

[11] No simple relation appears to exist between the rank of two sets and the ranks of their cross-cut and union. For example, let $M_1 = \{36, 54\}$, $M_2 = \{18, 24\}$. Then $[M_1, M_2] = M_1$, $(M_1, M_2) = M_2$ so that M_2 contains M_1. The rank of M_1 is 108, while the rank of M_2 is 72. Thus, the L. C. M. of the ranks of M_1 and M_2 is not the rank of their cross-cut, the G. C. D. of their ranks is not the G. C. D. of their union, nor does one rank divide the other.

Now assume that $(\mathfrak{A}, C_k) = \mathfrak{O}$. If (3.2) holds for $n \geqq n_0 > 0$ we have from the recursion relation

$$U_{n_0+k-1+\lambda} = C_1 U_{n_0+k-2+\lambda} + \cdots + C_k U_{n_0-1+\lambda}$$

$$U_{n_0+k-1} = C_1 U_{n_0+k-2} + \cdots + C_k U_{n_0-1}.$$

But by (3.2), $U_{n_0+k-r+\lambda} \equiv U_{n_0+k-r} \pmod{\mathfrak{A}}$, $(r = 1, 2, \cdots, k)$. Hence by subtraction

$$C_k(U_{n_0-1+\lambda} - U_{n_0-1}) \equiv 0 \qquad (\mathrm{mod}\ \mathfrak{A}).$$

Since $(C_k, \mathfrak{A}) = \mathfrak{O}$, (3.2) holds for $n \geqq n_0 - 1$. Therefore (3.2) holds for $n \geqq 0$ and (U) is purely periodic.

THEOREM 6.2.[12] *Let $\mathfrak{O}$ be a domain of integrity, K a finite extension of its quotient field. Let $\Phi = \Phi(x_1, \cdots, x_r; y)$ be a polynomial in the $r + 1$ indeterminates $x_1, \cdots, x_r, y$ with coefficients in K and let $\omega_1, \omega_2, \cdots, \omega_r$ be r fixed integers of K. Define a sequence (V) over K by*

$$V_n = \Phi(\omega_1^n, \omega_2^n, \cdots, \omega_r^n; n), \qquad (n = 0, 1, 2, \cdots).$$

Then (V) is linear, and satisfies a recursion of the form (3.2) with coefficients in $\mathfrak{O}$. Every sequence which is linear over $\mathfrak{O}$ may be thus obtained by a suitable choice of the extension field K and polynomial Φ.

PROOF. Let N be the maximum degree of Φ in the x, and M the maximum degree of Φ in y. Suppose that the polynomial Φ when written in the form

$$\Phi = \sum_{(k)} \Gamma_{(h)}(p) x_1^{k_1} x_2^{k_2} \cdots x_r^{k_r} \qquad 0 \leqq k_1 + k_2 + \cdots + k_r \leqq N$$

has precisely s terms $X_{(k)} = x_1^{k_1} \cdots x_r^{k_r}$. Here the $\Gamma_{(k)}(y)$ are polynomials in y of degree $\leqq M$ with coefficients in K. Let $\Omega_k = \omega_1^{k_1} \cdots \omega_r^{k_r}$. Then there exists a polynomial $f(x) = x^t - c_1 x^{t-1} - \cdots - c_t$ with coefficients in $\mathfrak{O}$ such that $f_{T-1}(\Omega_{(k)}) = 0$, $(k = 1, \cdots, s)$. Let

$$F(x) = f(x)^M = x^T - D_1 x - \cdots - D_T.$$

Then each $\Omega_{(k)}$ is a root of $F(x) = 0$ of multiplicity at least M, and $D_1, \cdots D_T$ lie in $\mathfrak{O}$. But

$$V_n = \sum_{h=1}^{s} \Gamma_{(h)}(n) \Omega_{(k)}^n. \tag{6.1}$$

Hence (V) satisfies the recurrence

$$Y_{n+T} = D_1 Y_{n+T-1} + \cdots + D_T Y_n \tag{6.2}$$

associated with $F(x)$.

Now let (V) be a linear sequence of $\mathfrak{O}$ defined by (6.2) and assume that the

[12] We discard the scheme of notation explained in section 2 for the statement and proof of this theorem.

distinct roots of the associated polynomial $F(x)$ are $\Omega_{(1)}, \cdots, \Omega_{(s)}$. The Ω then lie in a finite extension K of the quotient field of $\mathfrak{D}$. Furthermore let each Ω be of multiplicity $\leqq M$. Then every solution of (6.2) in $\mathfrak{D}$ is of the form (6.1) if the coefficients of the polynomials $\Gamma_{(k)}(n)$ are suitably chosen in K. It suffices then to take $\Phi = \sum_{k=1}^{s} \Gamma_{(h)}(y)X_{(k)}$ where $X_{(1)}, \cdots, X_{(s)}$ and y are now our indeterminates.

THEOREM 6.21. *If $\mathfrak{D}$ is a domain of integrity and (U), (V) are linear over $\mathfrak{D}$, then the sequence (UV) whose general term is U_nV_n is also linear over $\mathfrak{D}$.*

PROOF. It suffices to observe that the product of two polynomials Φ_1 and Φ_2 with different indeterminates x but the same y with coefficients in the finite extension fields K_1 and K_2 is again a polynomial of the same form whose coefficients lie in the union of K_1 and K_2.

We shall call (UV) the *product* of the sequences (U) and (V) writing $(U)\cdot(V) = (UV)$. The operation of obtaining (UV) from (U) and (V) will be called *multiplication* of sequences. If we define the sum of two sequences as in Ward [5] by $(U) + (V) = (U + V)$, the following theorem is evident.

THEOREM 6.3. *The set of all sequences linear over a domain of integrity forms a commutative ring with respect to the operations of addition and multiplication defined above.*

This ring contains as its unit element the sequence 1, 1, 1, $\cdots$ satisfying the recursion $Y_{n+1} = Y_n$. It in general has no finite basis and is not a domain of integrity.

We may apply the operation of multiplication to divisibility sequences. The following theorem has important applications (Ward [6]).

THEOREM 6.4. *The product $(U)\cdot(V)$ of two divisibility sequences (U) and (V) is again a divisibility sequence. The set of places of apparition of any prime divisor $\mathfrak{P}$ of $(U)\cdot(V)$ is the union of the sets of places of apparition of $\mathfrak{P}$ in (U) and (V). The ranks of apparition of $\mathfrak{P}$ in $(U)\cdot(V)$ are contained among the ranks of apparition of $\mathfrak{P}$ in (U) and (V).*

The proof is simple, and will be omitted here. It is essential that the ideal divisor be a prime.

V. LUCASIAN SEQUENCES

7. We lose little generality[13] by assuming that a given divisibility sequence is *normal*; that is, $U_0 = 0$, $U_1 = 1$. (Ward [1], section 11.) If in addition (U) is Lucasian, the results of part III and IV of the paper yield at once a great deal of information about the divisors of (U) and their places of apparition. For any ideal $\mathfrak{A}$ prime to C_k and such that the quotient ring $\mathfrak{D}/\mathfrak{A}$ is finite is at

[13] Since any number divides zero, every divisor of a divisibility sequence divides U_0. This fact restricts the divisors at the outset unless $U_0 = 0$. Since 1 divides every number, U_1 divides every term of (U). If U_1 is not zero, we can divide it out of the sequence obtaining a new divisibility sequence in which $U_1 = 1$.

once a divisor of (U), and all its ranks of apparition divide the restricted period of (U). The least common multiple r of these ranks of apparition gives us the period of the places of apparition of $\mathfrak{A}$ in the sequence. In contrast to the behavior of the period and restricted period of (U) modulo $\mathfrak{A}$, the rank r modulo $\mathfrak{A}$ does not appear to be effectively calculable merely from a knowledge of the ranks of constituent factors of $\mathfrak{A}$. However if $\mathfrak{A} = [\mathfrak{B}, \mathfrak{C}]$, the set of places of apparition of $\mathfrak{A}$ are effectively calculable from the places of apparition of $\mathfrak{B}$ and $\mathfrak{C}$ by virtue of theorems **4.0** and **5.3**.

I hope to discuss the theory of Lucasian sequences in relation to the operation of multiplication of sequences defined in part IV in some detail elsewhere.

CALIFORNIA INSTITUTE OF TECHNOLOGY

REFERENCES

G. BIRKHOFF, 1 Bulletin American Math. Soc. vol. 45 (1934) pp. 613-619.
R. D. CARMICHAEL, 1. Quarterly Journal of Mathematics, vol. 48 (1920) pp. 343-372.
L. E. DICKSON, 1. *History*, volume 1, chapter XVII.
MARSHALL HALL, 1. American Journal of Mathematics, vol. 47 (1936) pp. 577-584.
E. LUCAS, 1. American Journal of Mathematics, vol. 1 (1878) pp. 184-240, pp. 289-321.
2. *Theorie des Nombres*, Paris 1891.
O. ORE, 1. These Annals, (2) vol. 36 (1935) pp. 406-437.
B. L. VAN DER WAERDEN, *Modern Algebra*, vol. 2, Berlin 1931.
M. WARD, 1. These Annals, vol. 38 (1937) pp. 725-732.
2. Transactions of the American Math. Soc., vol. 33 (1931) pp. 162-164.
3. Bulletin of the American Math. Soc., vol. 42 (1936) pp. 843-845.
4. Transactions of the American Math. Soc., vol. 41 (1937), pp. 276-286.
5. Transactions of the American Math. Soc., vol. 35 (1933) pp. 600-628.
6. To appear in the Transactions of the American Math. Soc.

Annals of Mathematics
Vol. 39, No. 1, January, 1938

GROUP RINGS AND EXTENSIONS. I

By Marshall Hall

(Received May 25, 1937)

1. **Introduction: The problem of extension.** A group G is said to be an *extension of a group A by a group H* if A is a normal subgroup of G and $G/A = H$. There is always at least one extension of A by H, namely their direct product, but in general there will be other extensions. The problem of extension is to construct and investigate all extensions of a given A by a given H.

We may write

$$G = 1A + \bar{u}A + \bar{v}A \cdots + \bar{w}A, \tag{1.1}$$

where each coset $\bar{u}A$ (and in particular its representative $\bar{u}$) is associated by the homomorphism which maps G onto H with an element u of H. The identity will always be taken as the representative of A. O. Schreier[1] has shown that a knowledge of (a) the automorphisms $A \rightleftarrows \bar{u}^{-1}A\bar{u}$ or symbolically

$$A \rightleftarrows A^u, \tag{1.2}$$

and (b) the factor set of (u, v) in

$$\bar{u}\,\bar{v} = \overline{uv}(u, v) \tag{1.3}$$

is sufficient to determine an extension G uniquely. The automorphisms and factor set are not arbitrary, and he gives the conditions which they must satisfy. R. Baer[2] has investigated the problem of extension in some detail, treating in particular the case when A is abelian.

This paper investigates central extensions, that is extensions in which the factor set of (u, v) is in the center of A, and includes the case A abelian. The automorphisms must form a group homomorphic to H. It is shown that the conditions which the factor sets must satisfy are paraphrases of identities on the defining relations of H. And principally it shows that the problem of central extensions of A is equivalent to a problem on vector moduli (or invariant vector spaces) over the group ring of H. This leads to the determination of all central extensions of A by H. For the proof of the principal theorem it is necessary to include (§5) two theorems of some interest in themselves, one very special theorem on extensions closely related to Galois theory, and another on a sort of algebraic-topological closure in semi-simple algebras.

---

[1] O. Schreier "Über die Erweiterung von Gruppen I" Monatshefte fur Mathematik und Physik, vol. 34 (1926) pp. 165–180, in particular Theorem I, p. 168.

[2] R. Baer "Erweiterung von Gruppen und ihren Isomorphismen" Mathematische Zeitschrift, vol. 38 (1934) pp. 375–416.

As an application of this theory, a new criterion is given for direct factors. This includes a theorem of Burnside's[3] and the well known theorem on complete groups as special cases. Further applications will appear in part II of this paper.

2. **Central extensions.** The conditions (see reference to Schreier) which the automorphisms and factor set must satisfy are that

$$(\alpha^v)^w = (v, w)^{-1}\alpha^{vw}(v, w) \tag{2.1}$$

and

$$(uv, w)(u, v)^w = (u, vw)(v, w) \tag{2.2}$$

hold for all α of A and all u, v, w of H. Now if all factors (u, v) are in B, the center of A, (2.1) requires that the automorphisms $A \rightleftarrows A^u$ form a group homomorphic to H. Let us denote by χ a particular way of assigning to each element of H an automorphism of A, where the automorphisms as assigned form a group homomorphic to H. Furthermore χ shall be fixed throughout this paper. A central extension of A by H inducing automorphisms in A by the rule χ will be called, with Baer, an $H - \chi$ extension, or, when no ambiguity can arise, simply an extension. This settles the condition (2.1) and we need now consider only (2.2).

The automorphism A^u depends solely upon the choice of $\bar{u}$ modulo B, and so we may alter $\bar{u}$ by a factor from B without changing the rule χ. We shall require that a representative $\bar{u}$ be changed only by multiplication by a factor from B. This requirement is vacuous if A is abelian. There always exists at least one $H - \chi$ extension of A, namely that in which all factors (u, v) are the identity, for then (2.2) is certainly satisfied. In this case, from (1.3), the representatives of the cosets form a group isomorphic to H, and we shall say that G is the normal product of A by H.

If in an extension G we change the representatives of the cosets by putting $\bar{\bar{u}} = \bar{u}\alpha(u), \cdots, \bar{\bar{w}} = \bar{w}\alpha(w)$, where $\alpha(u), \cdots, \alpha(w)$ are elements from B, then we find

$$\bar{\bar{u}}\bar{\bar{v}} = \overline{\overline{uv}}\alpha(u)^v\alpha(v)\alpha(uv)^{-1}(u, v) \tag{2.3}$$

determining the new factor set:

$$(u, v)^* = \alpha(u)^v\alpha(v)\alpha(uv)^{-1}(u, v). \tag{2.4}$$

Since the extension G has not been altered, it is reasonable to consider factor sets so related as equivalent and we shall write

$$(u, v)^* \sim (u, v) \tag{2.5}$$

[3] Burnside "Theory of groups of finite order." Cambridge, 2nd edition 1911, p. 327.

whenever a function $\alpha(u)$ with domain H and range B exists satisfying (2.4). It is easily verified that this is a true equivalence, being reflexive, symmetric, and transitive.

If $(u, v)_1$ and $(u, v)_2$ are factor sets satisfying (2.2) and we define $(u, v)_3 = (u, v)_1(u, v)_2$, then it follows that $(u, v)_3$ also satisfy (2.2) and are a permissable factor set. In this definition of product for factor sets, there is an identity, the set in which all factors are the identity, and an inverse, the set in which (u, v) is replaced by $(u, v)^{-1}$; moreover if $(u, v)_1^* \sim (u, v)_1$ and $(u, v)_2^* \sim (u, v)_2$, then $(u, v)_1^*(u, v)_2^* \sim (u, v)_1(u, v)_2$. Hence the totality of all factor sets forms a group (necessarily abelian since $(u, v)_1(u, v)_2 = (u, v)_2(u, v)_1$) even if we identify all equivalent sets. The group in which equivalent sets are identified will be called the group of extensions.

If H is finite we define, following Nakayama[4]

$$f(v) = \prod_u (u, v). \tag{2.6}$$

Multiplying (2.2) over all u, we obtain

$$f(w)f(v)^w = f(vw)(v, w)^n, \tag{2.7}$$

where n is the order of H. On comparison with (2.4), we see that

$$(v, w)^n \sim 1. \tag{2.8}$$

Again if m is a multiple of the order of every element of B, since the (u, v) are elements of B

$$(v, w)^m = 1. \tag{2.9}$$

Hence the following theorem holds:

THEOREM 2.1: *The order of any element of the group of extensions divides the order of H and the least common multiple of orders of elements of B.*

COROLLARY: *If m and n are relatively prime, then the only $H - \chi$ extension of A by H is the normal product of A by H.*

3. **Operators and defining relations.** Let α be any element of B and let $\alpha \rightleftarrows \alpha^u$ under the automorphism of A induced by any element u of H. If n is any rational integer $(\alpha^n)^u = (\alpha^u)^n$ which we may denote by α^{nu}. Let us write $\alpha^{ru}\alpha^{sv} \cdots \alpha^{tw} = \alpha^{ru+sv\cdots+tw}$. In this way we may attach an unambiguous meaning to α^h where h is any element of the group ring H^* formed of sums of elements of H with rational integral coefficients. We shall speak

[4] T. Nakayama "Über die Beziehung zwischen den Faktorensystemen und der Normklassengruppe eines galoisschen Erweiterungskörper" Mathematische Annalen, vol. 112 (1935) pp. 85–91. The application of this "Japanese homomorphism" to groups is due to some unpublished work of R. Baer.

of H^* as the group ring of operators on B. The operators have the following properties:

$$(\alpha\beta)^h = \alpha^h\beta^h, \qquad \alpha^{h+k} = \alpha^h\alpha^k, \qquad (\alpha^h)^k = \alpha^{hk}. \tag{3.1}$$

The free abelian group on symbols

$$a^1, a^u, a^v, \cdots \qquad (1, u, v, \cdots \text{ from } H) \tag{3.2}$$

has among its automorphisms the permutations of its generators

$$w = \begin{pmatrix} a^u \\ a^{uw} \end{pmatrix}, \tag{3.3}$$

and these automorphisms form a group isomorphic to H. Its elements are of the form a^h with operators h from H^* and obey the rules (3.1). In addition, since the generators (3.2) are bound by no relation, the following implication holds

$$a^h = 1 \rightarrow h = 0. \tag{3.4}$$

Such a group or the direct product of such groups will be called *operator free*.

Let $x, y, z, \cdots$ be a set of generators of H and let each element of H be represented by a definite "word" in $x, y, z, \cdots$. Now let us choose the representatives of the cosets as the corresponding words on $\bar{x}, \bar{y}, \bar{z}, \cdots$. These generate a subgroup $\bar{G}$ of G. The free group F on $\mathbf{x}, \mathbf{y}, \mathbf{z}, \cdots$ is mapped homomorphically onto $\bar{G}$ by the correspondence $\mathbf{x} \rightarrow \bar{x}$, $\mathbf{y} \rightarrow \bar{y}$, etc., and in turn the correspondence $\bar{x} \rightarrow x, \bar{y} \rightarrow y$ etc. maps $\bar{G}$ homomorphically onto H. Hence we have $F \rightarrow \bar{G} \rightarrow H$, $\mathbf{x} \rightarrow \bar{x} \rightarrow x$. Let the defining relations of H be

$$\Phi_i(x, y, z, \cdots) = 1 \qquad i = 1, 2, \cdots, r. \tag{3.5}$$

Those words of F mapped onto the identity of H form[5] the least normal subgroup R of F containing $\Phi_i(\mathbf{x}, \mathbf{y}, \mathbf{z}, \cdots)$. The factor set

$$(u, v)\, \overline{uv}^{-1}\bar{u}\,\bar{v} \tag{3.6}$$

is mapped onto the identity of H, and so the corresponding words of F lie in R. These words in fact generate R[6] and since the factors are in B

$$R \rightarrow \bar{B} \rightarrow 1, \tag{3.7}$$

where $\bar{B}$ is a subgroup of B. Hence

$$\Phi_i(\bar{x}, \bar{y}, \bar{z}, \cdots) = \alpha_i \qquad i = 1, \cdots, r \tag{3.8}$$

with α_i in $\bar{B}$. Moreover (3.8) serves to determine the image in $\bar{B}$ of any element of R, since we need only calculate the automorphs of the α's (i.e. trans-

---

[5] Reidemeister "Einführung in die Kombinatorische Topologie" pp. 35–36.

[6] O. Schreier "Die Untergruppen der freien Gruppen" Abhandlungen der Hamburgischen Universität, vol. 5 (1927) pp. 161–183 esp. p. 173.

forms of the Φ's) and the products of these. In particular (3.8) determines the values of all the factor sets. Now since B is abelian, the commutator subgroup C of R is mapped onto the identity of $\bar{G}$.

$$(3.9) \qquad C \to 1 \to 1.$$

Hence elements of F congruent modulo C are mapped onto the same element of $\bar{G}$.

THEOREM 3.1: *A necessary and sufficient condition that* $\Phi_i(\bar{x}, \bar{y}, \bar{z}, \cdots) = \alpha_i$, $i = 1, \cdots, r$ *define a central extension of A by H is that whenever* $\prod_i \Phi_i^{h_i} = 1$ *in F modulo C then* $\prod_i \alpha_i^{h_i} = 1$.

The element $\prod_i \Phi_i^{h_i}$ in F modulo C is mapped into $\prod_i \Phi_i^{h_i}$ in $\bar{G}$. Hence from $\prod_i \Phi_i^{h_i} = 1$ follows $\prod_i \alpha_i^{h_i} = 1$. On the other hand, the conditions (2.2) which the α_i must satisfy to define a central extension are homomorphs of identities in F and *à* fortiori in F modulo C. For (2.2) is a paraphrase of

$$(3.10) \qquad (\overline{uvw}^{-1}\,\overline{uv}\,\bar{w})(\bar{w}^{-1}(\overline{uv}^{-1}\,\bar{u}\,\bar{v})\bar{w}) = (\overline{uvw}^{-1}\,\bar{u}\,\overline{vw})(\overline{vw}^{-1}\,\bar{v}\,\bar{w}).$$

An identity $\prod \Phi_i^{h_i} \equiv 1 \pmod{C}$ may be considered as belonging to a vector

$$(3.11) \qquad [h_1, h_2, \cdots, h_r]$$

of elements from H^*.

THEOREM 3.2: *The vectors* $[h_1, h_2, \cdots, h_r]$ *for which* $\prod_i \Phi_i^{h_i} \equiv 1 \pmod{C}$ *form a right modulus M over* H^*.

For

$$\prod_i \Phi_i^{h_i} \equiv 1, \qquad \prod_i \Phi_i^{k_i} \equiv 1 \to \prod_i \Phi_i^{h_i+k_i} \equiv 1 \qquad \pmod{C}$$

since R modulo C is abelian, whence

$$[h_1, h_2, \cdots, h_r], \qquad [k_1, k_2, \cdots, k_r] \in M \to [h_1 + k_1, \cdots, h_r + k_r] \in M.$$

Also

$$\prod_i \Phi_i^{h_i} \equiv 1, \qquad \lambda \in H^* \to \prod_i \Phi_i^{h_i\lambda} \equiv 1 \qquad \pmod{C},$$

or

$$[h_1, h_2, \cdots, h_r] \in M, \qquad \lambda \in H^* \to [h_1\lambda, h_2\lambda, \cdots, h_r\lambda] \in M.$$

The ring H^* consists of the element of H with integral coefficients and may be imbedded in an algebra $\mathfrak{A}$ consisting of elements of H with rational coefficients. There is a least vector modulus $\mathfrak{M}$ over $\mathfrak{A}$ which contains M. If every vector of $\mathfrak{M}$ with components in H^* is in M, we shall say that the modulus M is *closed*.

THEOREM 3.3: *The right modulus M of theorem* 3.2 *is closed.*

A vector X of $\mathfrak{M}$ is of the form $X_1\lambda_1 + \cdots + X_s\lambda_s$ with $X_1, \cdots, X_s$ from M and $\lambda_1, \cdots, \lambda_s$ from $\mathfrak{A}$. If m is a multiple of all the denominators of the λ's, then mX is a vector of M, since $m\lambda_1, \cdots, m\lambda_s$ are elements of H^*. Hence to prove that M is closed it is sufficient to show that $X = [h_1, \cdots, h_r]$ with h_i

from H^* is in M if mX in M, where m is a rational integer. Now $\prod_i \Phi_i^{h_i}$ is an element of R modulo C. R, a subgroup of the free group F, is a free group itself (see reference to Schreier above). Hence R modulo C is a free abelian group. Consequently

$$mX \epsilon M \rightarrow \prod_i \Phi_i^{h_i m} = \left(\prod_i \Phi_i^{h_i}\right)^m = 1 \rightarrow \prod_i \Phi_i^{h_i} = 1 \rightarrow X \epsilon M.$$

4. **Relations in the group ring.** In $\Phi_i(\bar{x}, \bar{y}, \cdots)$ when the representatives of the cosets are changed by putting $\bar{x} = \xi\bar{\bar{x}}$, $\bar{y} = \eta\bar{\bar{y}}$, $\cdots$ we have

$$\Phi_i(\bar{x}, \bar{y}, \cdots) = \Phi_i(\bar{\bar{x}}, \bar{\bar{y}}, \cdots)\xi^{x_i}\eta^{y_i}\cdots \tag{4.1}$$

if we move $\xi, \eta, \cdots$ to the end of the word by the rules $\xi^h\bar{w} = \bar{w}\xi^{hw}$ and $\xi^h\bar{w}^{-1} = \bar{w}^{-1}\xi^{hw^{-1}}$. The elements x_i, y_i, $\cdots$ of H^* obtained in this way play an important rôle in the following fundamental theorems, which taken with Theorems 3.1 determine all central extensions of A by H.

THEOREM 4.1: $\prod \Phi_i^{h_i} \equiv 1 \pmod{C}$ *if and only if* $\sum x_i h_i = 0$, $\sum y_i h_i = 0$, *etc.*

THEOREM 4.2: *The group of* $H - \chi$ *extensions of* A *is the group of*

$$[\alpha_1, \alpha_2, \cdots, \alpha_r],$$

with $\prod_i \alpha_i^{h_i} = 1$ *for all relations of Theorem* 3.1, *modulo* $[\beta_1, \beta_2, \cdots, \beta_r]$ *where* $\beta_i = \xi^{x_i}\eta^{y_i}\cdots$ $i = 1, \cdots, r$, $\xi, \eta, \cdots$ *arbitrary elements of* B.

The proof of Theorem 4.2 is easy. $[\alpha_1, \cdots, \alpha_r]$ defines an extension [by putting $\Phi_i(\bar{x}, \bar{y}, \cdots) = \alpha_i$] if and only if all relations $\prod \alpha_i^{h_i} = 1$ of Theorem 3.1 are satisfied. Since the factor sets generate $\bar{B}$, $\alpha_i = (u, v) \cdots (w, z)$, and if $(x, y)^{(1)}(x, y)^{(2)} = (x, y)^{(3)}$ then $\alpha_i^{(1)}\alpha_i^{(2)} = \alpha_i^{(3)}$. In the group of extensions those factor sets are equivalent to the identity which can be obtained from the identity by a change of the representatives of the cosets. But these yield precisely $\beta_i = \xi^{x_i}\eta^{y_i}\cdots$ $(i = 1, \cdots, r)$ with arbitrary $\xi, \eta, \cdots$ from B.

The proof of Theorem 4.1 is on the other hand very difficult. Let T denote the set of vectors in $H^*[t_1, t_2, \cdots, t_r]$ satisfying

$$\begin{aligned} x_1t_1 + x_2t_2 + \cdots + x_rt_r &= 0, \\ y_1t_1 + y_2t_2 + \cdots + y_rt_r &= 0. \end{aligned} \tag{4.2}$$

T is at once seen to be a right vector modulus over H^*.

Consider the set V_r of all vectors $[k_1, k_2, \cdots, k_r]$ with r components from H^*. If $A = [a_1, \cdots, a_r]$ and $B = [b_1, \cdots, b_r]$ are two vectors of V_r, and if

$$a_1b_1 + \cdots + a_rb_r = 0, \tag{4.3}$$

we shall say that A annihilates B on the left or that B annihilates A on the right. To use another term, geometrically suggestive, we may say that A is orthogonal to B on the left or that B is orthogonal to A on the right. It is evident that the vectors which annihilate any set S of vectors on the right

(left) form a right (left) modulus over H^*. Thus T is the right modulus of vectors annihilating $[x_1, \cdots, x_r]$, $[y_1, \cdots, y_r]$ etc. on the right.

Let A be the operator free abelian group whose elements are a^h with h from H^*, and let χ be the automorphisms (3.3). Then

$$\Phi_i(\bar{x}, \bar{y}, \cdots) = a^{u_i} \qquad i = 1, \cdots, r \tag{4.4}$$

defines an $H - \chi$ extension of A, by Theorem 3.1 if and only if

$$\prod (a^{u_i})^{h_i} = 1 \tag{4.5}$$

for every $[h_1, \cdots, h_r]$ of M. Since A is operator free (4.5) implies that

$$u_1 h_1 + \cdots + u_r h_r = 0. \tag{4.6}$$

Hence the vectors $U = [u_1, u_2, \cdots, u_r]$ associated with extensions (4.4) are those annihilating M on the left. Now we turn to the

Lemma: *The only extension of an operator free group $A = \{a^h\}$ by H is the normal product of A by H.* This lemma is the special case of Theorem 5.1 in which $Z = 1$ and is also proved separately in §5 by a different method. By the lemma, for any $[u_1, u_2, \cdots, u_r]$ of U $[a^{u_1}, \cdots, a^{u_r}] \sim [1, \cdots, 1]$. As in Theorem 4.2 $[\beta_1, \cdots, \beta_r] \sim [1, \cdots, 1]$ if and only if $\beta_i = \xi^{x_i}\eta^{y_i} \cdots$, $i = 1, \cdots, r$. Here $a^{u_i} = (a^h)^{x_i}(a^k)^{y_i} \cdots$ whence $u_i = hx_i + ky_i + \cdots$, $i = 1, \cdots, r$. Hence U is the left vector modulus over H^* with a basis $[x_1, \cdots, x_r]$, $[y_1, \cdots, y_r]$ etc. By definition T consists of the vectors annihilating on the right a basis of U (and consequently all of U).

We now know that U is the set of all vectors annihilating M on the left, and that T is the set of all vectors annihilating U on the right. From this it is a triviality that $M \subseteq T$. [This much could be proved easily even without the lemma.] But from this and the fact that M is a closed right modulus over H^* we shall be able to conclude that $M = T$, proving our theorem.

Here we turn to some very general considerations on rings.

Definition 4.1: *In a ring $\mathfrak{R}$, the right (left) annihilators of the left (right) annihilators of a set of vectors S shall be called the right (left) linear closure $\bar{S}_r$ (or $\bar{S}_l$) of S.*

Definition 4.2: *If every right (left) vector modulus of $\mathfrak{R}$ is its own right (left) linear closure, we shall say that $\mathfrak{R}$ is linearly closed.* We note that the closure of S always includes S, and that if $\mathfrak{R}$ is associative and distributive, then the closure of S always includes the least modulus including S.

In §5 we shall prove (Theorem 5.2) that a semi-simple algebra is linearly closed. If $\mathfrak{A}$ is the group algebra of H over the rational field, then $\mathfrak{A}$ is semi-simple.[7] Hence $\mathfrak{M}$, the least right vector modulus of $\mathfrak{A}$ containing M is its own linear closure. Let $\mathfrak{U}$ be the left modulus of vectors annihilating all vectors of $\mathfrak{M}$ on the left and $\mathfrak{M}'$ (= $\mathfrak{M}$ by the theorem) be the right modulus of vectors

[7] J. H. M. Wedderburn "Lectures on Matrices" American Mathematical Society Colloquium Publications, vol. 17 (1934) p. 168.

annihilating $\mathfrak{U}$ on the right. U is the intersection of $\mathfrak{U}$ with V_r. For any vector annihilating M on the left will also annihilate $\mathfrak{M}$ on the left [since any element of $\mathfrak{M}$ is equal to an element of M divided by a rational integer]. Similarly T is the intersection of $\mathfrak{M}'$ with V_r. But by the theorem on semi-simple algebras, $\mathfrak{M}' = \mathfrak{M}$ and so T is the intersection of $\mathfrak{M}$ with V_r. By Theorem 3.3 M is the intersection of $\mathfrak{M}$ with V_r. Hence $M = T$ and the proof of Theorem 4.1 is complete, except for the auxiliary theorems of §5.

As an example of the application of Theorems 4.1 and 4.2 let us suppose that H is a dihedral group defined by

$$x^n = 1, \qquad y^2 = 1, \qquad yx = x^{-1}y. \tag{4.7}$$

A central extension of a group A by H will be defined by

$$\bar{x}^n = \alpha, \qquad \bar{y}^2 = \beta, \qquad \bar{y}\,\bar{x}\,\bar{y}^{-1}\bar{x} = \gamma) \tag{4.8}$$

with α, β, γ from B, the center of A, and they must satisfy certain conditions

$$\alpha^{h_1}\beta^{h_2}\gamma^{h_3} = 1. \tag{4.9}$$

From Theorem 4.1 the equations which h_1, h_2, h_3 must satisfy are

$$\begin{aligned} (1 + x + \cdots + x^{n-1})h_1 + (y + x)h_3 &= 0, \\ (1 + y)h_2 + (1 - x)h_3 &= 0, \end{aligned} \tag{4.10}$$

and we find a basis of the right modulus $[h_1, h_2, h_3]$ to be

$$\begin{aligned} &[x - 1, 0, 0]; \quad [0, y - 1, 0], \\ &[y + 1, 0, -(1 + x + \cdots + x^{n-1})], \\ &[0, x - 1, 1 - x^{-1}y]. \end{aligned} \tag{4.11}$$

The conditions which α, β, γ must satisfy are therefore

$$\begin{aligned} &\alpha^{x-1} = 1; \quad \beta^{y-1} = 1, \\ &\alpha^{y+1}\gamma^{-(1+x+\cdots+x^{n-1})} = 1, \\ &\beta^{x-1}\gamma^{1-x^{-1}y} = 1. \end{aligned} \tag{4.12}$$

We know that these must be paraphrases of identities on $\mathbf{x}^n$, $\mathbf{y}^2$, $\mathbf{y}\mathbf{x}\mathbf{y}^{-1}\mathbf{x}$ and their transforms in F and these are in fact:

$$\begin{aligned} &\mathbf{x}^{-1}(\mathbf{x}^n)\mathbf{x}\cdot(\mathbf{x}^n)^{-1} = 1, \\ &\mathbf{y}^{-1}(\mathbf{y}^2)\mathbf{y}\cdot(\mathbf{y}^2)^{-1} = 1, \\ &(\mathbf{yxy}^{-1}\mathbf{x})\cdot\mathbf{x}^{-1}(\mathbf{yxy}^{-1}\mathbf{x})\mathbf{x}\cdot\mathbf{x}^{-2}(\mathbf{yxy}^{-1}\mathbf{x})\mathbf{x}^2 \cdots \mathbf{x}^{-n+1}(\mathbf{yxy}^{-1}\mathbf{x})\mathbf{x}^{n-1} \\ &\qquad\qquad = \mathbf{y}(\mathbf{x}^n)\mathbf{y}^{-1}\cdot\mathbf{x}^n, \\ &\mathbf{x}^{-1}(\mathbf{y}^2)\mathbf{x}\cdot[\mathbf{y}^{-1}\mathbf{x}(\mathbf{yxy}^{-1}\mathbf{x})\mathbf{x}^{-1}\mathbf{y}]^{-1}\cdot(\mathbf{y}^2)^{-1}\cdot(\mathbf{yxy}^{-1}\mathbf{x}) = 1, \end{aligned} \tag{4.13}$$

(left) form a right (left) modulus over H^*. Thus T is the right modulus of vectors annihilating $[x_1, \cdots, x_r]$, $[y_1, \cdots, y_r]$ etc. on the right.

Let A be the operator free abelian group whose elements are a^h with h from H^*, and let χ be the automorphisms (3.3). Then

$$\Phi_i(\bar{x}, \bar{y}, \cdots) = a^{u_i} \qquad i = 1, \cdots, r \tag{4.4}$$

defines an $H - \chi$ extension of A, by Theorem 3.1 if and only if

$$\prod (a^{u_i})^{h_i} = 1 \tag{4.5}$$

for every $[h_1, \cdots, h_r]$ of M. Since A is operator free (4.5) implies that

$$u_1h_1 + \cdots + u_rh_r = 0. \tag{4.6}$$

Hence the vectors $U = [u_1, u_2, \cdots, u_r]$ associated with extensions (4.4) are those annihilating M on the left. Now we turn to the

Lemma: *The only extension of an operator free group $A = \{a^h\}$ by H is the normal product of A by H.* This lemma is the special case of Theorem 5.1 in which $Z = 1$ and is also proved separately in §5 by a different method. By the lemma, for any $[u_1, u_2, \cdots, u_r]$ of U $[a^{u_1}, \cdots, a^{u_r}] \sim [1, \cdots, 1]$. As in Theorem 4.2 $[\beta_1, \cdots, \beta_r] \sim [1, \cdots, 1]$ if and only if $\beta_i = \xi^{x_i}\eta^{y_i} \cdots$, $i = 1, \cdots, r$. Here $a^{u_i} = (a^h)^{x_i}(a^k)^{y_i} \cdots$ whence $u_i = hx_i + ky_i + \cdots$, $i = 1, \cdots, r$. Hence U is the left vector modulus over H^* with a basis $[x_1, \cdots, x_r]$, $[y_1, \cdots, y_r]$ etc. By definition T consists of the vectors annihilating on the right a basis of U (and consequently all of U).

We now know that U is the set of all vectors annihilating M on the left, and that T is the set of all vectors annihilating U on the right. From this it is a triviality that $M \subseteq T$. [This much could be proved easily even without the lemma.] But from this and the fact that M is a closed right modulus over H^* we shall be able to conclude that $M = T$, proving our theorem.

Here we turn to some very general considerations on rings.

Definition 4.1: *In a ring $\mathfrak{R}$, the right (left) annihilators of the left (right) annihilators of a set of vectors S shall be called the right (left) linear closure $\bar{S}_r$ (or $\bar{S}_l$) of S.*

Definition 4.2: *If every right (left) vector modulus of $\mathfrak{R}$ is its own right (left) linear closure, we shall say that $\mathfrak{R}$ is linearly closed.* We note that the closure of S always includes S, and that if $\mathfrak{R}$ is associative and distributive, then the closure of S always includes the least modulus including S.

In §5 we shall prove (Theorem 5.2) that a semi-simple algebra is linearly closed. If $\mathfrak{A}$ is the group algebra of H over the rational field, then $\mathfrak{A}$ is semi-simple.[7] Hence $\mathfrak{M}$, the least right vector modulus of $\mathfrak{A}$ containing M is its own linear closure. Let $\mathfrak{U}$ be the left modulus of vectors annihilating all vectors of $\mathfrak{M}$ on the left and $\mathfrak{M}'$ ($= \mathfrak{M}$ by the theorem) be the right modulus of vectors

[7] J. H. M. Wedderburn "Lectures on Matrices" American Mathematical Society Colloquium Publications, vol. 17 (1934) p. 168.

annihilating $\mathfrak{U}$ on the right. U is the intersection of $\mathfrak{U}$ with V_r. For any vector annihilating M on the left will also annihilate $\mathfrak{M}$ on the left [since any element of $\mathfrak{M}$ is equal to an element of M divided by a rational integer]. Similarly T is the intersection of $\mathfrak{M}'$ with V_r. But by the theorem on semi-simple algebras, $\mathfrak{M}' = \mathfrak{M}$ and so T is the intersection of $\mathfrak{M}$ with V_r. By Theorem 3.3 M is the intersection of $\mathfrak{M}$ with V_r. Hence $M = T$ and the proof of Theorem 4.1 is complete, except for the auxiliary theorems of §5.

As an example of the application of Theorems 4.1 and 4.2 let us suppose that H is a dihedral group defined by

$$x^n = 1, \qquad y^2 = 1, \qquad yx = x^{-1}y. \tag{4.7}$$

A central extension of a group A by H will be defined by

$$\bar{x}^n = \alpha, \qquad \bar{y}^2 = \beta, \qquad \bar{y}\,\bar{x}\,\bar{y}^{-1}\bar{x} = \gamma) \tag{4.8}$$

with α, β, γ from B, the center of A, and they must satisfy certain conditions

$$\alpha^{h_1}\beta^{h_2}\gamma^{h_3} = 1. \tag{4.9}$$

From Theorem 4.1 the equations which h_1, h_2, h_3 must satisfy are

$$\begin{aligned} (1 + x + \cdots + x^{n-1})h_1 + (y + x)h_3 &= 0, \\ (1 + y)h_2 + (1 - x)h_3 &= 0, \end{aligned} \tag{4.10}$$

and we find a basis of the right modulus $[h_1, h_2, h_3]$ to be

$$\begin{aligned} &[x - 1, 0, 0]; \quad [0, y - 1, 0], \\ &[y + 1, 0, -(1 + x + \cdots + x^{n-1})], \\ &[0, x - 1, 1 - x^{-1}y]. \end{aligned} \tag{4.11}$$

The conditions which α, β, γ must satisfy are therefore

$$\begin{aligned} &\alpha^{x-1} = 1; \quad \beta^{y-1} = 1, \\ &\alpha^{y+1}\gamma^{-(1+x+\cdots+x^{n-1})} = 1, \\ &\beta^{x-1}\gamma^{1-x^{-1}y} = 1. \end{aligned} \tag{4.12}$$

We know that these must be paraphrases of identities on $\mathbf{x}^n$, $\mathbf{y}^2$, $\mathbf{yxy}^{-1}\mathbf{x}$ and their transforms in F and these are in fact:

$$\begin{aligned} &\mathbf{x}^{-1}(\mathbf{x}^n)\mathbf{x}\cdot(\mathbf{x}^n)^{-1} = 1, \\ &\mathbf{y}^{-1}(\mathbf{y}^2)\mathbf{y}\cdot(\mathbf{y}^2)^{-1} = 1, \\ &(\mathbf{yxy}^{-1}\mathbf{x})\cdot\mathbf{x}^{-1}(\mathbf{yxy}^{-1}\mathbf{x})\mathbf{x}\cdot\mathbf{x}^{-2}(\mathbf{yxy}^{-1}\mathbf{x})\mathbf{x}^2 \cdots \mathbf{x}^{-n+1}(\mathbf{yxy}^{-1}\mathbf{x})\mathbf{x}^{n-1} \\ &\qquad\qquad = \mathbf{y}(\mathbf{x}^n)\mathbf{y}^{-1}\cdot\mathbf{x}^n, \\ &\mathbf{x}^{-1}(\mathbf{y}^2)\mathbf{x}\cdot[\mathbf{y}^{-1}\mathbf{x}(\mathbf{yxy}^{-1}\mathbf{x})\mathbf{x}^{-1}\mathbf{y}]^{-1}\cdot(\mathbf{y}^2)^{-1}\cdot(\mathbf{yxy}^{-1}\mathbf{x}) = 1, \end{aligned} \tag{4.13}$$

5. **Two auxiliary theorems.**

THEOREM 5.1: *Suppose that* (1) $b_1, b_2, \cdots, b_s$ *are the generators of a free abelian group* A *and* (2) H *induces automorphmiss on* A *which permute* $b_1, \cdots, b_s$ *transitively. Then, if* Z *is the subgroup of* H *which leaves* b_1 *fixed and* $C(Z)$ *is its commutator subgroup, the group of extensions of* A *by* H *is isomorphic to* $Z/C(Z)$.[8]

From (2.7) we have for any factor set (u, v)

$$(u, v)^n = f(u)^v f(v) f(uv),^{-1} \tag{5.1}$$

where n is the order of H. Conversely if in A there are functions $Q(u, v)$ and $q(u)$ such that

$$Q(u, v)^n = q(u)^v q(v) q(uv)^{-1} \tag{5.2}$$

for all u, v of H, then the $Q(u, v)$ are a factor set. By (2.4) $Q(u, v)^n \sim 1$ form a factor set and hence satisfy (2.2). The n^{th} roots of these relations (unique since A is a free abelian group) are the same relations holding for $Q(u, v)$ and hence $Q(u, v)$ form a factor set. Again, since n^{th} roots are unique in A, $q(u)$ in (5.2) defines $Q(u, v)$ uniquely. Moreover if $q_1(u) \to Q_1(u, v)$ and $q_2(u) \to Q_2(u, v)$ then $q_1(u)q_2(u) \to Q_1(u, v)Q_2(u, v)$. Hence the group of extensions of A by H is the multiplicative group of q's for which $q(u)^v q(v) q(uv)^{-1}$ is an n^{th} power, making equivalent those q's defining equivalent Q's.

We shall first settle the question of equivalence. Let $q(y) \to Q(u, v)$ and $q'(u) \to Q'(u, v)$ and suppose $Q(u, v) \sim Q'(u, v)$. Then $Q'(u, v) = h(u)^v h(v) h(uv)^{-1} Q(u, v)$ whence

$$[h(u)^v h(v) h(uv)^{-1} Q(u, v)]^n = q'(u)^v q'(v) q'(uv)^{-1}, \tag{5.3}$$

and if we put, defining a function $a(u)$,

$$q'(u) = q(u) h(u)^n a(u), \tag{5.4}$$

we find that the condition on $a(u)$ is

$$a(u)^v a(v) = a(uv). \tag{5.5}$$

To solve (5.5) set $\prod_u a(u) = b$ and multiply over all u. Then

$$b^v [a(v)]^n = b \tag{5.6}$$

for all v. If $b = b_1^{t_1} \cdots b_s^{t_s}$ then

$$[a(v)]^n = b_1^{t_1 - t_1(v)} \cdots b_s^{t_s - t_s(v)}, \tag{5.7}$$

where $t_1(v), \cdots, t_s(v)$ is a permutation of $t_1, \cdots, t_s$ according to the permutation of the b's induced by v. Since the right hand side of (5.7) is an n^{th} power, in particular we have $t_1 \equiv t_1(v) \pmod n$. Since this congruence holds for all v

[8] This theorem has been known in one form or another for some years but has never been published. It has been attributed to some work on Artin's on Class Field Theory. Its present form is approximately that given by R. Baer in a series of lectures at the Institute for Advanced Study.

and since H induces a transitive permutation on the b's, $t_1 \equiv t_2 \equiv \cdots \equiv t_s$ (mod n) and so

$$b = (b_1 b_2 \cdots b_s)^{t_1} a^n, \tag{5.8}$$

whence

$$[a(v)]^n = b^{1-v} = a^{n-nv}, \tag{5.9}$$

and

$$a(v) = a^{1-v}. \tag{5.10}$$

Now (5.10) satisfies (5.5) whatever a may be and hence is the most general solution of (5.5). Consequently $q'(u) \sim q(u)$ if and only if

$$q'(u) = q(u)h(u)^n a^{1-u}, \tag{5.11}$$

and $h(u)$ and a may be chosen arbitrarily.

Let a decomposition of H into cosets of Z be

$$H = Z + Z\bar{w} + \cdots + Z\bar{t}, \tag{5.12}$$

where it is important that the representatives be written on the right, since Z need not be a normal subgroup of H. As $b_1^z = b_1$ for any z of Z, $b_1^u = b_1^{\bar{u}}$ where $\bar{u}$ is the representative of the coset in (5.12) to which u belongs. Hence $b_1, b_2, \cdots, b_s = b_1, b_1^{\bar{w}}, \cdots, b_1^{\bar{t}}$ respectively. To solve

$$q(u)^v q(v) \equiv q(uv) \qquad (\text{mod } A^n), \tag{5.13}$$

put $q(u) = b_1^{Q(u)}$ where

$$Q(u) = \sum_{\bar{w}} q(u, \bar{w})\bar{w}. \tag{5.14}$$

Then (5.13) is equivalent, because of the independence of $b_1, \cdots, b_1^{\bar{t}}$ to

$$q(u, \overline{\bar{w}v^{-1}}) + q(v, \bar{w}) \equiv q(uv, \bar{w}) \qquad (\text{mod } n) \tag{5.15}$$

for all u, v, w of H. For $v = 1$ we find

$$q(1, \bar{w}) \equiv 0 \qquad (\text{mod } n). \tag{5.16}$$

Put $\bar{w} = 1$ in (5.15) and we have

$$q(u, \overline{v^{-1}}) \equiv q(uv, 1) - q(v, 1), \tag{5.17}$$

determining all q's in terms of those whose second argument is the identity. Set $q(u, 1) = r(u)$ and we have from (5.17) since

$$\overline{(vz)^{-1}} = \overline{v^{-1}} \qquad (z \text{ from } Z),$$

$$q(u, \overline{v^{-1}}) \equiv r(uv) - r(v) \equiv r(uvz) - r(vz). \tag{5.18}$$

Now as $r(1) \equiv 0$ from (5.16) it follows, putting $v = 1$

$$r(uz) \equiv r(z) + r(u) \qquad (\text{mod } n). \tag{5.19}$$

Thus, for arguments within Z, $r(z)$ is a homomorphic mapping of Z onto the additive group of integers modulo n. Hence $r(z)$ is a character of Z made abelian, that is of $Z/C(Z)$, since n, the order of H, is a multiple of the order of Z.

Now we reverse the argument. Let $r(z)$ be any homomorphic mapping of Z onto the additive group of integers modulo n, that is, any character of $Z/C(Z)$. In addition let $r(\bar{w}^{-1})$, $\bar{w} \neq 1$, be an arbitrary residue modulo n, and set

$$r(u) \equiv r(z(u)) + r(\overline{u^{-1}}^{-1}), \tag{5.20}$$

where $u^{-1} = z(u)^{-1}\overline{u^{-1}}$. From this we verify that

$$r(uz) \equiv r(z) + r(u) \tag{5.21}$$

for any z of Z and any u of H. Again

$$r(uvz) - r(vz) \equiv r(z) + r(uv) - r(z) - r(v) \equiv r(uv) - r(v). \tag{5.22}$$

Hence $r(uv) - r(v)$ depends only on the coset to which v^{-1} belongs in (5.12) and we may define a function $q(u, \overline{v^{-1}})$ by

$$q(u, \overline{v^{-1}}) \equiv r(uv) - r(v). \tag{5.23}$$

Hence

$$\begin{aligned} q(u, \overline{\bar{w}\,v^{-1}}) - q(v, \bar{w}) &\equiv r(uv\bar{w}^{-1}) - r(v\bar{w}^{-1}) + r(v\bar{w}^{-1}) - r(\bar{w}^{-1}) \\ &\equiv r(uv\bar{w}^{-1}) - r(\bar{w}^{-1}) \equiv q(uv, \bar{w}). \end{aligned} \tag{5.24}$$

Thus the function $q(u, \bar{w})$ satisfies (5.15) and determines an extension of A by H, where

$$q(u) = b_1^{Q(u)}, \qquad Q(u) = \sum q(u, \bar{w})\bar{w}. \tag{5.25}$$

Here

$$\begin{aligned} Q(u) &\equiv \sum q(u, \bar{w})\bar{w} \equiv \sum [r(u\bar{w}^{-1}) - r(\bar{w}^{-1})]\,\bar{w} \\ &\equiv \sum r(z(u\bar{w}^{-1})\bar{w}) + \sum [r(\overline{\bar{w}u^{-1}}^{-1})\bar{w}u^{-1}]\,u - \sum r(\bar{w}^{-1})\bar{w} \\ &\equiv Q_1(u) + Q_2(u)u - Q_3. \end{aligned} \tag{5.26}$$

Put $b_1^{Q_3} = a$, where since $Q_3 = \sum r(\bar{w}^{-1})\bar{w}$, a does not depend on u. Now since $b_1^{\bar{w}u} b_1^{\overline{\bar{w}u}}$, it follows that $b_1^{Q_2(u)} = b_1^{Q_3} = a$. Hence

$$q(u) \equiv b_1^{Q_1(u)} a^{u-1} \pmod{A^n} \tag{5.27}$$

where $Q_1(u)$ depends solely upon the value of $r(u)$ for arguments in Z. By (5.11) it follows that

$$q(u) \sim b_1^{Q_1(u)}, \tag{5.28}$$

with $Q_1(u)$ depending solely upon a character of $Z/C(Z)$. Factors a^{u-1} and $h(u)^n$ will leave the character $r(z)$ unchanged and so to each character there is exactly one extension. Moreover the product of two extensions is derived from the product of the two corresponding characters. And since a finite abelian group is isomorphic to its group of characters, it follows that the group of extensions is isomorphic to $Z/C(Z)$, and the theorem is proved.

We now give a separate proof of the lemma needed for Theorem 4.1 although it is the special case of Theorem 5.1 in which $Z = 1$.

LEMMA. *The only extension of an operator free group $A = \{a^n\}$ by H is the normal product of A by H.*[9]

In A every element b is expressible uniquely as

$$b = \prod_t a^{b(t)t} = \prod_t (b; t)^t, \tag{5.29}$$

where $(b; t) = a^{b(t)}$. Hence for a factor set

$$(u, v) = \prod (u, v; t)^t, \tag{5.30}$$

and (2.2) becomes because of the independence of $a, a^u, \cdots, a^w$

$$(uv, w; t)(u, v; tw^{-1}) = (u, vw; t)(v, w; t) \tag{5.31}$$

for all u, v, w, t of H. If we now put $\bar{u} = \bar{u} \prod_t (u, t^{-1}; 1)^{-t}$ for all u of H, we may verify from direct calculation and substitution from (5.31) that

$$\bar{\bar{u}}\,\bar{\bar{v}} = \bar{\bar{uv}}. \tag{5.32}$$

Hence the new representatives form a group, or in other words the extension is the normal product of A by H.

We note in passing that the same proof will also hold when A is any operator free group. Theorem 5.1 may also be generalized to hold for the direct product of groups $A_1, A_2, \cdots$. The group of extensions will be isomorphic to the direct product of $Z_1/C(Z_1), Z_2/C(Z_2), \cdots$.

THEOREM 5.2: *A semi-simple algebra is linearly closed.*

A semi-simple algebra is by the theorem of Wedderburn[10] the direct sum of simple algebras, and each simple algebra is a complete matric algebra over a division algebra. Our proof of the theorem will consist in showing (1) that a division algebra is linearly closed; (2) That if a ring $\mathfrak{R}$ with unit is linearly closed, then $\mathfrak{R}_n$, the complete matric algebra of degree n over $\mathfrak{R}$, is linearly closed; and (3) that if rings $\mathfrak{R}^{(1)}, \mathfrak{R}^{(2)}, \cdots, \mathfrak{R}^{(h)}$ are linearly closed, then their direct sum is linearly closed.[11]

(1) *A division algebra is linearly closed.*

Let $\mathfrak{D}$ be a division algebra and let $X = \{[x_1, \cdots, x_r]\}$ be a left vector modulus over $\mathfrak{D}$. We must find the totality of vectors $Y = \{[y_1, \cdots, y_r]\}$ satisfying

$$x_1y_1 + \cdots + x_ry_r = 0 \tag{5.33}$$

---

[9] Compare A. Scholz "Ein Beitrag zur Theorie der Zusammensetzung endlicher Gruppen" Mathematische Zeitschrift, vol. 32 (1930) pp. 187–189. Scholz uses the Galois theory of fields in the proof of his theorem.

[10] J. H. M. Wedderburn "On hypercomplex numbers" Proceedings of the London Mathematical Society (2) vol. 6 (1908) pp. 77–118.

[11] The notion of a closure of a set S as $\bar{S}$, the annihilator of the annihilator of S, or as the set orthogonal to the set orthogonal to S, has appeared in various forms, and the analogue of Theorem 5.2 is always fundamental. See J. von Neumann "On Regular Rings" Proceedings of the National Academy of Sciences, vol. 22 (1936) pp. 707–713, especially Lemma 12. In topology see Pontrjagin "Theory of topological commutative groups" Annals of Mathematics (2) vol. 35 (1934) pp. 361–388, especially page 369.

for every vector of X, and the totality of vectors $\bar{X} = \{[\bar{x}_1, \cdots, \bar{x}_r]\}$ satisfying

$$\bar{x}_1 y_1 + \cdots + \bar{x}_r y_r = 0 \tag{5.34}$$

for every vector of Y. We must then prove $\bar{X} = X$. If every vector of X vanishes, then Y contains all vectors over D and in particular $[1, 0, \cdots, 0]$; $[0, 1, 0, \cdots, 0]$; $\cdots$; $[0, 0, \cdots, 1]$ whence from (5.34) every vector of $\bar{X}$ vanishes and $\bar{X} = X$. If X contains non vanishing vectors, then it contains one vector U in which some component, say x_1, is different from zero. Hence X also contains $x_1^{-1}U = X_1$, a vector whose first component is unity, $X_1 = [1, x_2, \cdots, x_r]$. Subtracting a suitable left multiple of X_1 from every vector of X, we have remaining a set of vectors $[0, x_2, \cdots, x_r]$. Unless these all vanish, there must be one in which some component, say x_2, is unity. Let $X_2 = [0, 1, x_3, \cdots, x_r]$ and subtract suitable left multiples of X_2 from these vectors, leaving vectors whose first two components vanish. Proceeding in this way, we find a basis $X_1, X_2, \cdots, X_s$ for X and the corresponding equations determining Y are

$$\begin{array}{lr} X_1: & 1\cdot y_1 + x_2^{(1)}y_2 + x_3^{(1)}y_3 + \cdots + x_r^{(1)}y_r = 0, \\ X_2: & 1\cdot y_2 + x_3^{(2)}y_3 + \cdots + x_r^{(2)}y_r = 0, \\ \cdots\cdot & \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ X_s: & 1\cdot y_s + x_{s+1}^{(s)}y_{s+1} + \cdots + x_r^{(s)}y_r = 0. \end{array} \tag{5.35}$$

Here we may calculate $y_s, y_{s-1}, \cdots, y_1$ successively in terms of $y_{s+1}, \cdots, y_r$. Hence components $y_{s+1}, \cdots, y_r$ are arbitrary in vectors Y and for each choice of these the remaining components are uniquely determined.

Now $\bar{X}$ is a left modulus and contains X. Hence to prove $\bar{X} = X$, we must show that any vector of $\bar{X}$ is a vector of X. Let $\bar{X}_1$ be a vector of $\bar{X}$. Then $\bar{X}_1 = \lambda_1 X_1 + \cdots + \lambda_s X_s + [0, 0, \cdots, \bar{x}_{s+1}, \cdots, \bar{x}_r]$ if $\lambda_1, \cdots, \lambda_s$ are suitably chosen. Hence $[0, 0, \cdots, 0, \bar{x}_{s+1}, \cdots, \bar{x}_r]$ is also a vector of $\bar{X}$, and as an annihilator of Y must satisfy

$$\bar{x}_{s+1}y_{s+1} + \cdots + \bar{x}_r y_r = 0 \tag{5.36}$$

for any vector $[y_1, \cdots, y_r]$ of Y. But in Y the components $y_{s+1}, \cdots, y_r$ are arbitrary and hence from (5.36) $\bar{x}_{s+1} = \cdots = \bar{x}_r = 0$ whence

$$\bar{X}_1 = \lambda_1 X_1 + \cdots + \lambda_s X_s$$

and $\bar{X}_1$ is a vector of X. Hence $\bar{X} = X$ and X is its own left closure. This argument may be duplicated interchanging "left" and "right." Hence every right (left) vector modulus over $\mathfrak{D}$ is its own right (left) linear closure and $\mathfrak{D}$ is linearly closed.

(2) *If a ring $\mathfrak{R}$ with unit is linearly closed, then $\mathfrak{R}_n$, the complete matrix algebra of degree n over $\mathfrak{R}$, is linearly closed.*

Consider a left vector modulus over $\mathfrak{R}_n$, $X = \{[X_1, \cdots, X_r]\}$. With each vector $[X_1, \cdots, X_r]$ in X let us associate a vector

$$[a_1, \cdots, a_n; a_{n+1}, \cdots, a_{2n}; \cdots; a_{(r-1)n+1}, \cdots a_{rn}]$$

in $\mathfrak{R}$, where $a_1, \cdots, a_n$ in order are the elements of the first row of X_1, $a_{n+1}, \cdots, a_{2n}$ in order are the elements of the first row of X_2, and so forth. It is easily seen that the totality of vectors $A = \{[a_1, \cdots, a_{rn}]\}$ arising in this way from X forms a left vector modulus over $\mathfrak{R}$.

Let $B(k, k) = (b_{ij})$ be the matrix of $\mathfrak{R}_n$ in which $b_{kk} = 1$ $b_{ij} = 0$ otherwise. Then $B(k, k)[X_1, \cdots, X_r] = [X_{1,k}, \cdots, X_{r,k}]$ where the k^{th} row of $X_{i,k}$ is the same as the k^{th} row of X_i and the rest of its rows vanish. If $C(m, k) = (c_{ij})$ where $c_{mk} = 1$, $c_{ij} = 0$ otherwise, then $C(m, k)[X_{1,k}, \cdots, X_{r,k}] = [U_{1,m}, \cdots, U_{r,m}] = U(m)$ where the m^{th} row of $U_{i,m}$ is the same as the k^{th} row of $X_{i,k}$ and the rest of the rows vanish. Taking $m = 1$, we see that A is also the totality of k^{th} rows from X. Any vector of X is the sum of n U's, that is, any vector of X can be formed from n appropriate vectors of A. Conversely any matrix vector $[X_1, \cdots, X_r]$ formed in this way from n arbitrary vectors of A is in X.

The annihilators $Y = \{[Y_1, \cdots, Y_r]\}$ of X are defined by

$$X_1Y_1 + \cdots + X_rY_r = 0, \tag{5.37}$$

which is equivalent to

$$\sum_{k,t} x_{ik}^t y_{kj}^t = 0 \qquad i, j = 1, \cdots, n, \tag{5.38}$$

where $X_t = (x_{ij}^t)$, $Y_t = (y_{ij}^t)$. But the x's in (5.38) are the vectors of A and (5.38) shows that Y is formed of columns chosen arbitrarily from vectors of B, the right annihilator of A. Similarly $\bar{X}$, the left annihilator of Y is formed of rows chosen arbitrarily from $\bar{A}$ the left annihilator of B. But as $\mathfrak{R}$ is linearly closed, $\bar{A} = A$ whence $\bar{X} = X$. Similarly any right vector modulus over $\mathfrak{R}_n$ is its own right closure. Hence $\mathfrak{R}_n$ is linearly closed.

(3) *If $\mathfrak{R}^{(1)}, \mathfrak{R}^{(2)}, \cdots, \mathfrak{R}^{(h)}$ are rings which are linearly closed, then their direct sum $\mathfrak{R} = \mathfrak{R}^{(1)} + \cdots + \mathfrak{R}^{(h)}$ is linearly closed.*

Consider a homogeneous linear equation in $\mathfrak{R}$

$$x_1y_1 + \cdots + x_ry_r = 0. \tag{5.39}$$

Now let

$$\begin{aligned} x_i &= x_i^{(1)} + \cdots + x_i^{(h)}, \\ y_i &= y_i^{(1)} + \cdots + y_i^{(h)}, \end{aligned} \tag{5.40}$$

with $x_i^{(1)}, y_i^{(1)}$ in $\mathfrak{R}^{(1)}$. Then since $\mathfrak{R}^{(i)}\mathfrak{R}^{(j)} = 0$ if $i \neq j$, (5.39) becomes

$$\begin{aligned} &x_1^{(1)}y_1^{(1)} + \cdots + x_r^{(1)}y_r^{(1)}, \\ &+ x_1^{(2)}y_1^{(2)} + \cdots + x_r^{(2)}y_r^{(2)}, \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ &+ x_1^{(h)}y_1^{(h)} + \cdots + x_r^{(h)}y_r^{(h)} = 0. \end{aligned} \tag{5.41}$$

And since $u^{(1)} + \cdots + u^{(h)} = 0$ with $u^{(t)}$ in $\mathfrak{R}^{(t)}$ if and only if $u^{(1)} = u^{(2)} = \cdots = u^{(h)} = 0$, (5.41) involves

$$x_1^{(t)}y_1^{(t)} + \cdots + x_r^{(t)}y_r^{(t)} = 0. \qquad t = 1, \cdots, h \tag{5.42}$$

Hence (5.39) is the equivalent of the family of equations (5.42) one in each of $\mathfrak{R}^{(1)}, \cdots, \mathfrak{R}^{(h)}$ respectively and $\mathfrak{R}$ is linearly closed if and only if its direct summands are all linearly closed.

6. A criterion for direct factors.

THEOREM 6.1: *If a group G has a subgroup A such that*

(1) *A is normal*

(2) *G induces no outer automorphisms on A*

(3) *the order of the center of A is prime to the index of A in G*

then G has a subgroup H such that $G = A \times H$.

PROOF: Write

$$G = 1 \cdot A + \bar{x}A + \bar{y}A + \cdots. \tag{6.1}$$

Since G induces no outer automorphisms on A, $A \rightleftarrows \bar{x}^{-1}A\bar{x}$ is an inner automorphism of A and there exists an element α of A such that

$$\bar{x}^{-1}A\bar{x} = \alpha^{-1}A\alpha \tag{6.2}$$

elementwise. Hence $\bar{x}\alpha^{-1}$ permutes with every element of A. We may change the representative of the coset to $\bar{\bar{x}} = \bar{x}\alpha^{-1}$. Let us suppose this done in (6.1), all the representatives $\bar{x}, \bar{y}, \cdots$ permuting with every element of A. The factors $(x, y) = \overline{xy}^{-1}\bar{x}\,\bar{y}$ will also permute with every element of A, and must be in B, the center of A. Hence G is a central extension of A by a group H where the automorphisms χ are all the identical automorphism. By Theorem 2.1 the order of an element of the group of extensions divides the order of H and the order of B. By assumption these numbers are relatively prime and hence from the corollary to Theorem 2.1 G is the normal product of A by H. The elements of H induce the identical automorphism on A, is permute with every element of A, and so

$$G = A \times H. \tag{6.3}$$

Theorem 6.1 includes the Theorem of Burnside (loc. cit.) as the special case in which A is a Sylow subgroup. Another special case is the theorem that a normal subgroup which is a complete group is a direct factor. For all the automorphisms of a complete group are inner automorphisms and the center of a complete group is the identity.

THE INSTITUTE FOR ADVANCED STUDY.

Annals of Mathematics
Vol. 39, No. 1, January, 1938

ON FINITE SYSTEMS OF LINEAR DIFFERENTIAL EQUATIONS OF INFINITE ORDER WITH CONSTANT COEFFICIENTS

By Gaines B. Lang

(Received June 5, 1937)

1. Introduction

This paper treats of systems using certain operators involving derivatives and generalized differences, the operators being, in general, of infinite order and having constant coefficients. A characteristic property of the operators considered is that they do not increase the exponential type[1] of functions on which they operate. Only solutions of exponential type not exceeding a fixed finite number are considered, and all such solutions are found. For differential operators these results have previously been obtained by Lettenmeyer[2] by quite different methods. Martin[3] considered a system of difference equations of finite order and real span with more general coefficients and with less stringent conditions on known functions and solutions.

The method used is essentially the Pincherle transformation[4] as adapted by Carmichael.[5]

I. The Single Differential Equation

2. Definitions

Let

$$(1) \qquad \beta(t) = \sum_{\nu=0}^{\infty} \beta_\nu t^\nu$$

be analytic for $|t| \leqq q$ where $q > 0$. Then there exists a positive constant ϵ such that $\beta(t)$ is analytic for $|t| \leqq q + \epsilon$. If $y(x)$ is a function of exponential type not exceeding q, then the series

$$\sum_{\nu=0}^{\infty} \beta_\nu y^{(\nu)}(x)$$

[1] A function $f(x)$ is of exponential type q if $\lim \sup_{n\to\infty} |f^{(n)}(x)|^{1/n} = q$. For an exposition of the characteristic properties of functions of exponential type see Carmichael: *Functions of Exponential Type*, Bulletin American Mathematical Society, April (1934) pp. 241-261.

[2] Lettenmeyer: *Systeme linearer Differentialgleichungen unendlich hoher Ordnung mit Polynomes beschränkten Grades als Koeffizienten.* Habilitationschrift zur Erlangung der Venia legendi. München 1927.

[3] W. T. Martin: *Linear Difference Equations with Arbitrary Real Spans.* Acta Mathematica. (1938).

[4] Pincherle: *Memoria della R. Academia della Scienza dell' Istituto di Bologna,* S IV, T IX, (1888) pp. 45-71, and elsewhere.

[5] Carmichael: *Systems of Linear Difference Equations and Expansions in Series of Exponential Functions.* Trans. Am. Math. Soc., vol. 35, pp. 1-28. The author also wishes to express gratitude to Professor Carmichael for direction of the present paper.

converges uniformly as to x in any finite region, and defines a function of exponential type not exceeding q. We define

$$\beta[y(x)] = \sum_{\nu=0}^{\infty} \beta_\nu y^{(\nu)}(x)$$

and shall call the function $\beta(t)$ the characteristic function of the operator β. It is clear that β is a linear operator. Also it is easily verified that if two such operators are added, the characteristic function of the sum equals the sum of the characteristic functions, and an analogous statement applies to multiplication. Hence a polynomial (and in particular a determinant) of operators of this class is defined (when multiplying a function of exponential type not exceeding q).

3. The Single Differential Equation

Now suppose $\phi(x)$ is a function of exponential type not exceeding q. Let

$$\phi(x) = \sum_{\nu=0}^{\infty} \frac{s_\nu x^\nu}{\nu!}. \tag{2}$$

Define

$$\psi(t) = \sum_{\nu=0}^{\infty} s_\nu / t^{\nu+1}. \tag{3}$$

Since $\phi(x)$ is of exponential type not exceeding q, the series defining $\psi(t)$ converges for $|t| > q$. Then

$$\phi(x) = \frac{1}{2\pi i} \int_c e^{xt} \psi(t)\, dt \tag{4}$$

where c is the circle $|t| = q + \epsilon$ where $\epsilon > 0$.

R. D. Carmichael[1] has pointed out that there exists a one to one correspondence between functions of exponential type and functions analytic at a point and vanishing there. For simplicity we take the point to be infinity.

We shall need the following lemma:

LEMMA: *If $\phi(x)$ is of exponential type not exceeding q, q finite, and if*

$$\phi(x) = \frac{1}{2\pi i} \int_c f(t)\, e^{xt}\, dt$$

where c is the circle $|t| = q + \epsilon$ where $\epsilon > 0$ and where $f(t)$ is analytic upon c, then the necessary and sufficient condition that $\phi(x) \equiv 0$ is that $f(t)$ be analytic within c.

That the condition is sufficient is obvious from the Cauchy-Goursat Theorem. Its necessity is easily shown by writing the integrand as a Laurent series with non-vanishing coefficients about a supposed singularity inside c and seeing that $\phi(x)$ cannot vanish identically in this case.

Now consider the equation

$$(5)\qquad \beta[y(x)] = \sum_{\nu=0}^{\infty} \beta_\nu y^{(\nu)}(x) = \phi(x)$$

where $\phi(x)$ is of exponential type not exceeding q, and $\beta(t)$ is analytic for $|t| \leqq q$. We seek the most general solution $y(x)$ of exponential type not exceeding q.

Let

$$(6)\qquad y(x) = \sum_{\kappa=0}^{\infty} \frac{y_\kappa x^\kappa}{\kappa!}$$

be a solution of (5) of exponential type not exceeding q if such exists. Define

$$(7)\qquad Y(t) = \sum_{\kappa=0}^{\infty} y_\kappa / t^{\kappa+1}.$$

The series in (7) converges for $|t| > q$. Then

$$(8)\qquad y(x) = \frac{1}{2\pi i} \int_c Y(t)\, e^{xt}\, dt$$

where c is the circle $|t| = q + \epsilon$, where $\epsilon > 0$ is so chosen that $\beta(t)$ does not vanish in the circular ring $q < |t| \leqq q + \epsilon$.

We may write

$$(9)\qquad y(x) = \frac{1}{2\pi i} \int_c \frac{Y(t)\, \beta(t)\, e^{xt}\, dt}{\beta(t)}.$$

Now both $Y(t)$ and $\beta(t)$ are analytic in a circular ring inclosing c in its interior. Hence the product $Y(t)\beta(t)$ possess a Laurent expansion (valid along c) obtained by multiplying the two power series (7) and (1) term by term. The coefficient of $1/t^{k+1}$, $\kappa \geqq 0$, is seen to be s_κ, since $y(x)$ is by hypothesis a solution of (5). Hence

$$(10)\quad y(x) = \frac{1}{2\pi i} \int_c \frac{\psi(t) + \theta(t)}{\beta(t)} e^{xt}\, dt = \frac{1}{2\pi i} \int_c \frac{\psi(t)}{\beta(t)} e^{xt}\, dt + \frac{1}{2\pi i} \int_c \frac{\theta(t)}{\beta(t)} e^{xt}\, dt$$

where $\theta(t)$ is analytic throughout c and its interior.

Since the numerator and the denominator of the second integrand of (10) are analytic, the only singularities arise from zeros of $\beta(t)$. We may write $1/\beta(t)$ as the sum of the principal parts of the Laurent expansion about each zero plus a power series. The second term of (10) can then be reduced to the form

$$\frac{1}{2\pi i} \int_c \frac{P_{\sigma-1}(t)\, e^{xt}\, dt}{\pi(t)}$$

where $\pi(t)$ is that polynomial of degree σ, where σ is the number of zeros of $\beta(t)$ within and upon c (multiple zeros counted multiply) such that $\pi(t)/\beta(t)$

is analytic within and upon c, and $P_{\sigma-1}(t)$ is a polynomial with arbitrary coefficients of degree $\sigma - 1$. (If $\sigma = 0$, then $P_{\sigma-1}(t) \equiv 0$.) Hence any solution of (5), of exponential type not exceeding q, must be in the form

$$(11) \qquad \begin{aligned} y(x) &= \frac{1}{2\pi i}\int_c \frac{\psi(t)}{\beta(t)} e^{xt}\,dt + \frac{1}{2\pi i}\int_c \frac{P_{\sigma-1}(t)}{\pi(t)} e^{xt}\,dt \\ &= \eta(x) + \rho(x). \end{aligned}$$

It is easily verified that $\eta(x)$ is of exponential type not exceeding q, and that it satisfies (5); also that $\rho(x)$ is of exponential type not exceeding q, and that it satisfies the homogeneous equation (12).

$$(12) \qquad \beta[\rho(x)] = \sum_{\nu=0}^{\infty} \beta_\nu \rho^{(\nu)}(x) = 0.$$

It is easily seen that under the hypothesis that $\beta(t)$ is analytic for $|t| \leqq q$ the exponential type of $\beta_\nu y^{(\nu)}(x)$ cannot exceed the exponential type of $y(x)$. Hence we have the theorem:

THEOREM 1. *In the equation*

$$(5) \qquad \sum_{\nu=0}^{\infty} \beta_\nu y^{(\nu)}(x) = \phi(x)$$

let $\phi(x)$ be a given function of exponential type not exceeding q, and let the β_ν be constants such that the function $\beta(t) = \sum_{\nu=0}^{\infty} \beta_\nu t^\nu$ is analytic for $|t| \leqq q$. Let

$$\phi(x) = \sum_{\nu=0}^{\infty} \frac{s_\nu x^\nu}{\nu!}$$

and define

$$\psi(t) = \sum_{\nu=0}^{\infty} s_\nu / t^{\nu+1}.$$

If σ is the number of zeros of $\beta(t)$ within and upon the circle $|t| = q$, define $\pi(t)$ as the polynomial of degree σ, with leading coefficient one, such that $\beta(t)/\pi(t)$ is different from zero within and upon the circle $|t| = q$. Let $P_{\sigma-1}(t)$ be defined as a polynomial of degree $\sigma - 1$ with arbitrary coefficients. Define $P_{-1}(t) \equiv 0$. Choose c as a circle with center at 0, radius $q + \epsilon$, where ϵ is such a positive number that $\beta(t)$ is analytic within and upon c, and does not vanish in the circular ring $q < |t| \leqq q + \epsilon$. Then the general solution of (5) of exponential type not exceeding q can be written in the form

$$y(x) = \frac{1}{2\pi i}\int_c \frac{\psi(t)}{\beta(t)} e^{xt}\,dt + \frac{1}{2\pi i}\int_c \frac{P_{\sigma-1}(t)}{\pi(t)} e^{xt}\,dt,$$

involving just σ arbitrary constants. Furthermore, if $\phi(x)$ is of exponential type q, then $y(x)$ is of exponential type q.

II. The Non-Singular Differential System

4. A Particular Solution

We consider now the system of equations

$$\sum_{j=1}^{n} \alpha_{ij}[g_j(x)] = \sum_{j=1}^{n} \sum_{\nu=0}^{\infty} \alpha_{ij\nu} g_j^{(\nu)}(x) = \phi_i(x), \qquad i = 1, \cdots, n, \tag{13}$$

where $\phi_i(x)$, $i = 1, \cdots, n$, are functions of exponential type not exceeding q, and the $\alpha_{ij\nu}$ are constants subject to the conditions that

$$\alpha_{ij}(t) = \sum_{\nu=0}^{\infty} \alpha_{ij\nu} t^{\nu} \qquad i, j = 1, \cdots, n,$$

are analytic for $|t| \leqq q$. If the determinant

$$|\alpha_{ij}(t)| = \Delta(t)$$

vanishes identically in t, the system (13) is said to be singular, otherwise non-singular. In this section we consider only the non-singular system.

We define the operator α_{ij} as follows

$$\alpha_{ij}[f(x)] = \sum_{\nu=0}^{\infty} \alpha_{ij\nu} f^{(\nu)}(x) \qquad i, j = 1, \cdots, n.$$

The operator Δ is defined as that combination of linear differential operators α_{ij} resulting by the ordinary expansion of $|\alpha_{ij}|$. Its characteristic function is easily seen to be $\Delta(t)$. We denote by A_{ij} the co-factor of α_{ij} in Δ, and by $A_{ij}(t)$ the co-factor of $\alpha_{ij}(t)$ in $\Delta(t)$.

In system (13) we multiply the i^{th} equation by A_{ik}, $i = 1, \cdots, n$, and sum, and obtain

$$\Delta g_k(x) = \sum_{i=1}^{n} A_{ik}[\phi_i(x)] \qquad k = 1, \cdots, n. \tag{14}$$

Each of the equations (14) is of the form (5), and from the preceding section has a particular solution of the form

$$g_k^*(x) = \frac{1}{2\pi i} \int_c \frac{1}{\Delta(t)} \left\{ \sum_{i=1}^{n} A_{ik}(t)\, \psi_i(t) \right\} e^{xt}\, dt, \qquad k = 1, \cdots, n$$

where

$$\psi_i(t) = \sum_{\nu=0}^{\infty} \frac{1}{t^{\nu+1}} \phi^{(\nu)}(0) \qquad i = 1, \cdots, n,$$

and c is a circle with center at 0 and radius $q + \epsilon$ where ϵ is a positive constant so chosen that each $\alpha_{ij}(t)$ is analytic on c and $\Delta(t)$ has no zero in the circular ring $q < |t| \leqq q + \epsilon$. Hence we have the theorem:

THEOREM 2. *In the system of equations*

$$\sum_{j=1}^{n} \sum_{\nu=0}^{\infty} \alpha_{ij\nu} g_j^{(\nu)}(x) = \phi_i(x), \qquad i = 1, \cdots, n, \tag{13}$$

let the $\phi_i(x)$ *be given functions of exponential type not exceeding* q, *and let the constants* $\alpha_{ij\nu}$ *be such that the functions*

$$\alpha_{ij}(t) = \sum_{\nu=0}^{\infty} \alpha_{ij\nu} t^{\nu}, \qquad i, j = 1, \cdots, n,$$

are analytic for $|t| \leqq q$. *Denote by* $\Delta(t)$ *the determinant* $|\alpha_{ij}(t)|$ *and suppose* $\Delta(t) \not\equiv 0$. *Let*

$$\phi_i(x) = \sum_{\nu=0}^{\infty} \frac{s_{i\nu} x^{\nu}}{\nu!} \qquad i = 1, \cdots, n,$$

and define

$$\psi_i(t) = \sum_{\nu=0}^{\infty} s_{i\nu}/t^{\nu+1} \qquad i = 1, \cdots, n.$$

Let c *be a circle of radius* $q + \epsilon$, *center at origin, where* $\epsilon > 0$ *is so chosen that each* $\alpha_{ij}(t)$ *is analytic upon and inside* c, *and such that* $\Delta(t)$ *has no zero in the circular ring* $q < (t) \leqq q + \epsilon$. *Denote by* $A_{ij}(t)$ *the co-factor of* $\alpha_{ij}(t)$ *in* $\Delta(t)$. *Then a particular solution of* (13) *of exponential type not exceeding* q *is*

$$g_k^*(x) = \frac{1}{2\pi i} \int_c \left\{ \sum_{i=1}^{n} A_{ik}(t)\, \psi_i(t) \right\} e^{xt}\, dt, \qquad k = 1, \cdots, n. \tag{15}$$

The general solution of (13) *of exponential type not exceeding* q *can be obtained by adding to* $g_k^*(x)$ *the general solution of the required character, of the homogeneous system*

$$\sum_{j=1}^{n} \sum_{\nu=0}^{\infty} \alpha_{ij\nu} \bar{g}_j^{(\nu)}(x) = 0 \qquad i = 1, \cdots, n. \tag{16}$$

Furthermore, if at least one of the functions $\phi_i(x)$ *is of exponential type precisely* q, *then at least one of the functions* $g_k^*(x)$ *is of exponential type precisely* q.

III. THE HOMOGENEOUS NON-SINGULAR SYSTEM

5. The General Solution

We see from the preceding section that in order to obtain the general solution of exponential type not exceeding q of (13) we need to obtain the general solution of this character of (16).

$$\sum_{j=1}^{n} \alpha_{ij}[g_j(x)] = \sum_{j=1}^{n} \sum_{\nu=0}^{\infty} \alpha_{ij\nu} \bar{g}_j^{(\nu)}(x) = 0, \qquad i = 1, \cdots, n. \tag{16}$$

We consider the slightly more general system

$$\sum_{j=1}^{r} \alpha_{ij}[\bar{g}_j(x)] = \sum_{j=1}^{r} \sum_{\nu=0}^{\infty} \alpha_{ij\nu}\bar{g}_j^{(\nu)}(x) = 0, \qquad i = 1, \cdots, n \tag{17}$$

where $r \leqq n$, and where the matrix $M(t) = (\alpha_{ij}(t))$ is of rank r as a function of t i.e., at least one r-rowed determinant of $M(t)$ does not vanish identically in t. A point is considered a λ-fold common zero of the r-rowed determinants of $M(t)$ if (1) Each r-rowed determinant of $M(t)$ vanishes there to an order λ or higher and (2) At least one r-rowed determinant of $M(t)$ vanishes there to the order of λ precisely.

The following theorem is proved by induction on r, using the results of II and the obvious analogy to the method of solving the corresponding algebraic system.

THEOREM 3. *In the system*

$$\sum_{j=1}^{r} \sum_{\nu=0}^{\infty} \alpha_{ij\nu}\bar{g}_j^{(\nu)}(x) = 0 \qquad i = 1, \cdots, n, \tag{17}$$

let the constants $\alpha_{ij\nu}$ be such that the functions

$$\alpha_{ij}(t) = \sum_{\nu=0}^{\infty} \alpha_{ij\nu}t^\nu, \qquad i = 1, \cdots, n, \qquad j = 1, \cdots, r,$$

are analytic for $|t| \leqq q$. Denote by $M(t)$ the matrix

$$(\alpha_{ij}(t)) \qquad i = 1, \cdots, n; j = 1, \cdots, r.$$

Let M be of rank r as a function of t, i.e., some r-rowed determinant of M does not vanish identically in t. Then the number of linearly independent solutions of (17) *of exponential type not exceeding q will be precisely σ, where σ is the number of common zeros of all the r-rowed determinants of $M(t)$. Further, each solution of* (17) *is in the form*

$$\bar{g}_k(x) = \sum_{\nu=1}^{m} \left\{ p_{k\nu 0} + \cdots + p_{k\nu\lambda_\nu-1} \frac{x^{\lambda_\nu-1}}{(\lambda_\nu - 1)!} \right\} e^{\rho_\nu x}, \qquad k = 1, \cdots, r, \tag{18}$$

where $\rho_1, \cdots, \rho_m$, are the distinct common zeros, of orders of multiplicities $\lambda_1, \cdots, \lambda_m$, respectively, of the r-rowed determinants of M.

From Theorem 3 it follows immediately that the following two theorems are true.

THEOREM 4. *The system*

$$\sum_{j=1}^{n} \sum_{\nu=0}^{\infty} \alpha_{ij\nu}\bar{g}_j^{(\nu)}(x) = 0, \qquad i = 1, \cdots, n \tag{19}$$

in which the constants $\alpha_{ij\nu}$ are such that the functions

$$\alpha_{ij}(t) = \sum \alpha_{ij\nu}t^\nu, \qquad i, j = 1, \cdots, n,$$

are analytic for $|t| \leqq q$, *and in which the determinant* $\Delta(t) = |\alpha_{ij}(t)|$ *is not identically zero, has* σ *linearly independent solutions of exponential type not exceeding* q, *where* σ *is the number of zeros of* $\Delta(t)$ *within and upon the circle* $(t) = q$, *multiple zeros counted multiply. If the distinct zeros of* $\Delta(t)$ *in this region are* $\rho_1, \cdots, \rho_m$, *of orders of multiplicities* $\lambda_1, \cdots, \lambda_m$, *respectively, then each solution of exponential type not exceeding* q *is in the form*

$$(18)\qquad \bar{g}_k(x) = \sum_{\nu=1}^{m}\left\{p_{k\nu 0} + \cdots + p_{k\nu\lambda_\nu-1}\frac{x^{\lambda_\nu-1}}{(\lambda_\nu-1)!}\right\}e^{\rho_\nu x}, \qquad k = 1, \cdots, n.$$

THEOREM 5. *For the system* (13), *in which the hypotheses of Theorem* 2 *are satisfied, any solution of exponential type not exceeding* q *can be written in the form*

$$(20)\qquad g_k(x) = \frac{1}{2\pi i}\int_c G_k(t)\, e^{xt}\, dt, \qquad k = 1, \cdots, n$$

where the functions $G_k(t)$ *are subject to the condition*

$$(21)\qquad \sum_{k=1}^{n} \alpha_{ik}(t)\, G_k(t) = \psi_i(t) + Q_i(t) \qquad i = 1, \cdots, n,$$

where c *and the* $\psi_i(t)$ *are defined as in Theorem* 2, *and the* $Q_i(t)$ *are analytic for* $|t| \leqq q$. *Conversely, any set of functions of the form* (20) *such that* (21) *is satisfied, will satisfy* (13).

IV. THE SINGULAR DIFFERENTIAL SYSTEM

6. The Singular System

We consider the system

$$(22)\qquad \sum_{j=1}^{n}\sum_{\nu=0}^{\infty} \alpha_{ij\nu}\, g_j^{(\nu)}(x) = \phi_i(x), \qquad i = 1, \cdots, n,$$

where the determinant $D(t) = |\alpha_{ij}(t)|$ is of rank $r < n$ as a function of t, i.e., every $(r+1)$-rowed determinant, but not every r-rowed determinant, of D vanishes identically in t. The method is an extension of those already used, and the development proceeds along the line suggested by the corresponding algebraic system. The results are given by the following two theorems:

THEOREM 6. *In the system*

$$(23)\qquad \sum_{j=1}^{n} \alpha_{ij}[g_j(x)] = \phi_i(x), \qquad i = 1, \cdots, n,$$

let the $\phi_i(x)$ *be functions of exponential type not exceeding* q, q *finite, and let the determinant* $D(t) = |\alpha_{ij}(t)|$ *be of rank* r *as a function of* t, *i.e., every* $(r+1)$-*rowed determinant of* D, *but not every* r-*rowed determinant of* D, *vanishes identically in* t. *Arrange the determinant so that* $\Delta(t)$, *the* r-*rowed determinant in the upper left hand corner of* D, *does not vanish identically. Then to* $g_j(x), j = r+1, \cdots, n$,

may be assigned the values of any arbitrary functions of exponential type not exceeding q, and the system (23) *reduces to*

$$\sum_{j=1}^{r} \alpha_{ij}[\bar{g}_j(x)] = \bar{\phi}_i(x), \qquad i = 1, \cdots, n, \tag{24}$$

where

$$\bar{\phi}_i(x) = \frac{1}{2\pi i}\int_c \frac{1}{\Delta(t)} \begin{vmatrix} \alpha_{11}(t) & \cdots & \alpha_{1r}(t) & \psi_1(t) \\ \cdots & \cdots & \cdots & \cdots \\ \alpha_{r1}(t) & \cdots & \alpha_{rr}(t) & \psi_r(t) \\ \alpha_{i1}(t) & \cdots & \alpha_{ir}(t) & \psi_i(t) \end{vmatrix} e^{xt}\, dt, \qquad i = 1, \cdots, n.$$

Any solution of (23) *is of the form*

$$g_k(x) = g_k^*(x) + \bar{g}_k(x) \tag{25}$$

where $\bar{g}_k(x)$ *is a solution of* (24) *of exponential type not exceeding q, and*

$$g_k^*(x) = \frac{1}{2\pi i}\int_c \frac{1}{\Delta(t)}\left\{\sum_{i=1}^{r} A_{ik}(t)\,\psi_i(t) - \sum_{i=1}^{r}\sum_{j=r+1}^{n} \alpha_{ij}(t)\,A_{ik}(t)\,G_j(t)\right\} e^{xt}\, dt,$$
$$k = 1, \cdots, n$$

where

$$G_j(t) = \sum_{\nu=0}^{\infty} \frac{g_j^{(\nu)}(0)}{t^{\nu+1}}.$$

Conversely, any function of the form (25) *is of exponential type not exceeding q, and is a solution of* (23).

THEOREM 7. *A necessary and sufficient condition for the existence of a solution, of exponential type not exceeding q, of the system*

$$\sum_{j=1}^{r} \alpha_{ij}[g_j(x)] = \bar{\phi}_i(x), \qquad i = 1, \cdots, n,$$

in which

$$\phi_i(x) = 0 \qquad i = 1, \cdots, r,$$

$$\phi_i(x) = \sum_{\nu=1}^{m}\left\{l_{i\nu 0} + \cdots + l_{i\nu\lambda_\nu - 1}\frac{x^{\lambda_\nu - 1}}{(\lambda_\nu - 1)!}\right\} e^{\rho_\nu x} = \sum_{\nu=1}^{m} \bar{\phi}_{i\nu}(x), \quad i = r+1, \cdots, n,$$

and in which the $\alpha_{ij}(t)$ *are analytic for* $|t| \leqq q$ *and in which*

$$\Delta(t) = |\alpha_{ij}(t)|_{1,r} \not\equiv 0$$

and $\Delta(t)$ *vanishes to the order* λ_ν *at* $t = \rho_\nu$ *is that* (1) *the rank of the augmented matrix is r and* (2) *if for* $t = \rho_\nu$ *the rank of the matrix is* $s_\nu < r$, *the rank of the augmented matrix of the system obtained by replacing* $\bar{\phi}_i(x)$ *by* $\bar{\phi}_{i\nu}(x)$ *is* s_ν.

V. Difference and Mixed Systems

7. Generalized Differences of Finite Order

The preceding method is readily adapted to the solution of a class of generalized difference and mixed systems of which the most general is

$$\sum_{j=1}^{n} \sum_{\nu=0}^{\infty} \sum_{\mu=0}^{\infty} c_{ij\mu\nu} g_j^{(\nu)}(x + a_{ij\mu}) = \phi_i(x), \qquad i = 1, \cdots, n,$$

where the $\phi_i(x)$ are functions of exponential type not exceeding q, and the $c_{ij\mu\nu}$ and $a_{ij\mu}$ are constants subject to certain restrictions. By methods analogous to those already used the following theorems are readily established.

Theorem 8. *The system*

$$\sum_{j=1}^{n} \sum_{\mu=0}^{\sigma} \sum_{\nu=0}^{k} c_{ij\mu\nu} g_j^{(\nu)}(x + a_{ij\mu}) = \phi_i(x), \qquad i = 1, \cdots, n,$$

in which the functions $\phi_i(x)$ are of exponential type not exceeding q, q finite, is equivalent, as regards solutions of exponential type not exceeding q, to the system of differential equations of infinite order

$$\sum_{j=1}^{n} \alpha_{ij}[g_j(x)] = \phi_i(x), \qquad i = 1, \cdots, n,$$

in which

$$\alpha_{ij}(t) = \sum_{\mu=0}^{\sigma} \sum_{\nu=0}^{k} c_{ij\mu\nu} t^{\nu} e^{a_{ij\mu} t} = \sum_{\nu=0}^{\infty} \alpha_{ij\nu} t^{\nu}, \qquad i, j = 1, \cdots, n.$$

Theorem 9. *The system*

$$\sum_{j=1}^{n} \sum_{\mu=0}^{\sigma} \sum_{\nu=0}^{\infty} c_{ij\mu\nu} g_j^{(\nu)}(x + a_{ij\mu}) = \phi_i(x), \qquad i = 1, \cdots, n,$$

in which the $\phi_i(x)$ are functions of exponential type not exceeding q, and in which the functions

$$r_{ij\mu}(t) = \sum_{\nu=0}^{\infty} c_{ij\mu\nu} t^{\nu} \qquad i, j = 1, \cdots, n, \quad \mu = 0, \cdots, \sigma,$$

are analytic for $|t| \leqq q$, is equivalent, as regards solutions of exponential type not exceeding q, to the system

$$\sum_{j=1}^{n} \alpha_{ij}[g_j(x)] = \phi_i(x), \qquad i = 1, \cdots, n,$$

in which

$$\alpha_{ij}(t) = \sum_{\mu=0}^{\sigma} r_{ij\mu} e^{a_{ij\mu} t} = \sum_{\nu=0}^{\infty} \alpha_{ij\nu} t^{\nu}, \qquad i, j = 1, \cdots, n.$$

8. Generalized Differences of Infinite Order

THEOREM 10. *In the equation*

$$\sum_{\mu=0}^{\infty} c_\mu g(x + a_\mu) = \phi(x) \tag{26}$$

let $\phi(x)$ be a known function of exponential type not exceeding q. Suppose

$$\phi(x) = \sum_{\nu=0}^{\infty} \frac{s_\nu x^\nu}{\nu!}.$$

Let there exist a $Q > q$ such that the series

$$\sum_{\mu=0}^{\infty} c_\mu e^{a_\mu t}$$

converges uniformly for $|t| \leqq Q$. Define

$$\beta(t) = \sum_{\mu=0}^{\infty} c_\mu e^{a_\mu t}.$$

Then the general solution of (26) of exponential type not exceeding q, is given by

$$g(x) = \frac{1}{2\pi i}\int_c \frac{\psi(t)}{\beta(t)}\, e^{xt}\, dt + \frac{1}{2\pi i}\int_c \frac{P_{\sigma-1}(t)}{\pi(t)}\, e^{xt}\, dt,$$

where

$$\psi(t) = \sum_{\nu=0}^{\infty} s_\nu / t^{\nu+1},$$

c is a circle about 0 with radius R such that $q < R < Q$ and such that $\beta(t) \neq 0$ for $q < |t| \leqq R$, $\pi(t)$ is that polynomial of smallest degree with leading coefficient unity such that $\beta^{(t)}/\pi(t)$ is analytic and different from zero for $|t| \leqq q$, and $P_{\sigma-1}(t)$ is a polynomial of degree $\sigma - 1$ with arbitrary coefficients, where σ is the degree of $\pi(t)$ (and where $P_{-1}(t) \equiv 0$). Further, if $\phi(x)$ is of exponential type precisely q, then $g(x)$ is of exponential type q.

THEOREM 11. *The system of difference equations of infinite order*

$$\sum_{j=1}^{n} \sum_{\mu=0}^{\infty} c_{ij\mu} g_j(x + a_{ij\mu}) = \phi_i(x), \qquad i = 1, \cdots, n,$$

in which there exists a $Q > q$ such that the series

$$\sum_{\mu=0}^{\infty} c_{ij\mu} e^{a_{ij\mu} t}, \qquad i, j = 1, \cdots, n$$

all converge uniformly for $|t| \leqq Q$, and in which the functions $\phi_i(x)$ are of exponential type not exceeding q, is equivalent, as regards solutions of exponential type not exceeding q, to the system of differential equations of infinite order

$$\sum_{j=1}^{n} \sum_{\nu=0}^{\infty} \alpha_{ij\nu} g_j^{(\nu)}(x) = \phi_i(x), \qquad i = 1, \cdots, n,$$

in which

$$\alpha_{ij}(t) = \sum_{\mu=0}^{\infty} c_{ij\mu} e^{a_{ij\mu}t} = \sum_{\nu=0}^{\infty} \alpha_{ij\nu} t^{\nu}, \qquad i, j = 1, \cdots, n.$$

THEOREM 12. *The system*

$$\sum_{j=1}^{n} \sum_{\mu=0}^{\infty} \sum_{\nu=0}^{\infty} c_{ij\mu\nu} g_j^{(\nu)}(x + a_{ij\mu}) = \phi_i(x), \qquad i = 1, \cdots, n,$$

in which the $\phi_i(x)$ are functions of exponential type not exceeding q, and in which there exists a $Q > q$ such that the double series

$$\sum_{\mu=0}^{\infty} \sum_{\nu=0}^{\infty} c_{ij\mu\nu} e^{a_{ij\mu}t}, \qquad i, j = 1, \cdots, n,$$

converge uniformly for $|t| \leqq Q$, is equivalent, for solutions of exponential type not exceeding q, to the system of differential equations of infinite order

$$\sum_{j=1}^{n} \sum_{\nu=0}^{\infty} \alpha_{ij\nu} g_j^{(\nu)}(x) = \phi_i(x), \qquad i = 1, \cdots, n,$$

where

$$\alpha_{ij}(t) = \sum_{\mu=0}^{\infty} \sum_{\nu=0}^{\infty} c_{ij\mu\nu} t^{\nu} e^{a_{ij\mu}t} = \sum_{\nu=0}^{\infty} \alpha_{ij\nu} t^{\nu}, \qquad i, j = 1, \cdots, n.$$

WEST GEORGIA COLLEGE,
GENOLA, GA.

Annals of Mathematics
Vol. 39, No. 1, January, 1938

IRREGULARITY IN COMPLEXES

By S. Wylie

(Received July 2, 1937)

Introduction. In a complex K_n a simplex σ_p is said to be regular if its star has the homology characters of an n-cell, the star being defined as the set of all simplexes of K having σ_p as a face. If all the simplexes are regular K is said to be a manifold. This paper results from investigations of irregularity and of questions concerning the extent to which the irregularity of a simplex implies the irregularity of its faces. Answers are given to two main problems around which interest has crystallized:

(i) Can K have all its simplexes regular except some of dimension p?

(ii) Can K have all its simplexes regular except some of dimension $\geqq p$?

The second question arises because the answer to the first is in the negative unless $p = 0$ or $n - 1$, and it can itself be answered in the affirmative; this can be seen by constructing an actual example for $p = n - 1$, and by further subdividing it to introduce irregularity in lower dimensions.

Lastly we make use of the technique of the earlier part to solve a problem left open in Tucker's thesis.[1] He finds three conditions on a cell-complex that it should be a manifold,—closure, star, and intercept uniformity; but it is not there proved that intercept uniformity is independent of the other two. We construct an example of a complex having star and closure uniformity but lacking intercept uniformity, and thereby establish the required independence.

The terminology of the paper is largely that of Lefschetz;[2] however, the term 'complement' is used instead of 'link complex'.

I am indebted to Professor Tucker for the suggestion of the problems treated here, and also for some of the ideas involved in their solution.

The complement of a simplex. *The complement of σ_p in K is defined to be the set of all simplexes τ_q such that $\sigma_p \tau_q$ is a simplex of K.* If K is closed,—and that is the case we will be considering,—this is a subcomplex of K. We will write this as K/σ_p.[3] It is well known that the star of σ_p in K has the characters of a cell, if and only if the complement, K/σ_p, has the characters of an $(n - p - 1)$-sphere; it is this form of the condition of regularity that we will use.

We will need the following lemma:

---

[1] A. W. Tucker, An Abstract Approach To Manifolds, Annals of Mathematics, Vol. 34, 1933.

[2] S. Lefschetz, Topology—American Mathematical Society Colloquium Publications.

[3] This notation is due to M. H. A. Newman, in "Intersection Complexes," Proceedings of the Cambridge Philosophical Society of October 1931.

LEMMA 1. *The complement of $\sigma_p\sigma_q$ is precisely the complement of σ_q in the subcomplex described as the complement of σ_p in K. Or, written in symbols,* $K/\sigma_p\sigma_q = (K/\sigma_p)/\sigma_q$.

This is obvious from the definition of the complement.

The first problem. *Can K have all its simplexes of dimension $\neq p$ regular, and still have some irregular p-simplexes?* Suppose that it can and that $p > 0$, and consider a σ_{p-1}, which is a face of an irregular σ_p; let $\sigma_p = A\sigma_{p-1}$. Since σ_{p-1} is regular, K/σ_{p-1} has the characters of an $(n - p)$-sphere; but further, K/σ_{p-1} is a complex whose only irregular simplexes are vertices. This follows from Lemma 1: if τ_q is an irregular simplex of K/σ_{p-1}, $\sigma_{p-1}\tau_q$ is an irregular simplex of K; consequently from our original hypothesis, K/σ_{p-1} can have only O-dimensional irregular simplexes, and it does have at least one such, namely A.

If, then, K exists with the required property, it is possible to find complexes having the stated properties of K/σ_{p-1},—namely to be spherelike (have the characters of a sphere) and to have irregular vertices but no further irregularities. Call such a complex M_r, and we will prove that $r = 1$. Let I_α, $\alpha = 1$ to k, be the irregular vertices, and let I denote the whole set; we will assume M to be subdivided sufficiently finely for no vertex to be joined by a 1-simplex to more than one I_α. Then let N_α be the simplicial neighbourhood, or star, of I_α, and L_α its boundary.

Now from Tucker's thesis (loc. cit. 1) we take the observation that, if K_1 and K_2 are complementary sets, closed and open, of an r-spherelike complex,

$$R_p(K_1) = R_{p+1}(K_2) + \delta_0^p - \delta_{r-1}^p .$$

We apply this to M, using first $M - N$ and N, and secondly $N + L$ and M mod $(N + L)$, where $N = \sum N_\alpha$ and $L = \sum L_\alpha$; we also use the fact that, since M mod $(N + L)$ is a manifold $R_p(M \bmod N + L) = R_{r-p}(M - N)$. These give us that

$$\begin{aligned} R_p(N + L) &= R_{p+1}(M \bmod N + L) + \delta_0^p - \delta_{r-1}^p \\ &= R_{r-p-1}(M - N) + \delta_0^p - \delta_{r-1}^p \\ &= R_{r-p}(N). \end{aligned}$$

But $N + L$ is the closed simplicial neighbourhood of I and consequently has the homology characters of I, which in this case can be expressed as

$$R_p(N + L) = R_p(I) = k\delta_0^p .$$

A corresponding result obtains for the torsion groups of N, giving us that

$$R_{r-p}(N) = k\delta_0^p ;\ \theta_{r-p}(N) = 0.$$

But we also know that $R_q(N) = R_{q-1}(L) - k\delta_1^q$, because N is a collection of stars of vertices, and L the corresponding collection of complements. Therefore

$$R_q(L) = k\delta_0^q + R_{q+1}(N)$$

$$= k\delta_0^q + k\delta_{r-1}^q$$

and $\theta_q^i(L) = 0$.

But $R_q(L) = \sum_\alpha R_q(L_\alpha)$; if $r - 1 \neq 0$, we know that $R_0(L) = k$, and so $R_0(L_\alpha) = 1$. Because L_α is itself a manifold we deduce that $R_{r-1}(L_\alpha) = 1$ and hence

$$R_q(L_\alpha) = \delta_0^q + \delta_{r-1}^q$$

$$\theta_q^i(L_\alpha) = 0.$$

In other words L_α is spherelike and I_α is regular, contrary to hypothesis. Consequently we have reached a contradiction unless $r = 1$. Or, for $p > 0$, Kn cannot have only irregular σ_p's unless $p = n - 1$.

The answer to the first problem is, then, that such a complex cannot exist unless $p = 0$ or $n - 1$. If $p = 0$, it is clear that we can construct such a complex; simply identify any pair of vertices of an n-dimensional manifold. The question now remains,—can we construct an n-dimensional complex all of whose simplexes of dimension $< n - 1$ are regular, and containing some irregular σ_{n-1}'s?

The second problem. Before constructing the complex described at the end of the last paragraph we will show that the solution to the second problem will follow at once from that construction. This problem is,—*Can we construct a K having all its simplexes of dimension $< p$ regular, which has irregular σ_p's?* We will call a complex having this property $(p - 1)$-semiregular. Suppose then that we have constructed our required $(n - 2)$-semiregular complex; we get a p-semiregular complex from it by making a simplicial subdivision arising from placing a new vertex on a σ_{n-p-2} which is a face of some irregular σ_{n-1}. For the lowest-dimensional simplex introduced by this subdivision, which lies inside an irregular simplex of K, is precisely of dimension $p + 1$.

Consequently we shall have given a complete answer to both of the problems proposed, if we can construct an $(n - 2)$-semiregular K_n.

Construction of an $(n - 2)$-semiregular K_n.

LEMMA 2. *A necessary and sufficient condition that K_n be $(n - 2)$-semiregular is that the complements of all the vertices be regular or $(n - 3)$-semiregular and be spherelike.*

This follows from the definitions, with the help of Lemma 1. Lemma 2 leads us to expect an inductive construction and proof for the required example, and so we will aim at finding an $(n - 2)$-semiregular K_n which also has the characters of a sphere.

Consider any σ_{n-2} which is a face of an irregular σ_{n-1} in such a K; to each irregular vertex of K/σ_{n-2} corresponds an irregular σ_{n-1} incident on it. Since K/σ_{n-2} is 1-spherelike, the simplest example will arise from the case where it is just a circle containing a vertex A, and a line segment AB. In this case σ_{n-2} is incident on only two irregular σ_{n-1}'s, namely $A\sigma_{n-2}$ and $B\sigma_{n-2}$. Further we can observe that $A\sigma_{n-2}$ and $B\sigma_{n-2}$ have unlike complements,—three points and one point; so that for our simplest example we will have to construct the set of irregular σ_{n-1}'s in such a way that only two are incident on any σ_{n-2}; and so that we can colour these σ_{n-1}'s in two colours, and have no two of the same colour incident along a σ_{n-2};—say, those whose complements are three points are black and the others white.

One such an $(n-1)$-complex can be described as the regular polytope which is the analogue of an octahedron; it is the set of all simplexes $C_1C_2 \cdots C_n$, where C is a variable running over the two values A and B. We can colour those $(n-1)$-simplexes having an even number of B's black, and the rest white.

Such a set is to form the irregular set of our K_n; it is an $(n-1)$-sphere, and we will consider it to bound a simple n-cell. We call such a configuration, where the boundary $(n-1)$-sphere is subdivided as an octahedral polytope, an O_n. Our example will simply be an O_n with certain boundary simplexes matched with simplexes interior to the cell.

To describe this matching we use the given colouring of the σ_{n-1}'s on the boundary, together with a colouring of the boundary σ_{n-2}'s. Any such σ_{n-2} can be written as $C_1C_2 \cdots C_{p-1}C_{p+1} \cdots C_n$, where again C can stand for A or B. We colour this σ_{n-2} black if the number of B's in $C_1 \cdots C_{p-1}$ is odd.

We now match each black σ_{n-1} with some σ'_{n-1} interior to the cell; and we require that all the white faces of σ_{n-1} shall be faces of its corresponding σ'_{n-1} but that otherwise the boundary of σ'_{n-1} shall be interior to O_n. In fact all the black simplexes of the boundary get matched to simplexes interior to O_n, and none of the white simplexes.

We further require that the σ'_{n-1}'s shall all be distinct and shall have distinct boundaries in the interior. The point set described in this way admits of a simplicial subdivision in such a way that the irregular set, originally the boundary of O_n, is subdivided precisely as it was in O_n. In this case, whatever the subdivision of the interior, the complex will be called a Q_n. We now assert that Q_n is $(n-2)$-semiregular.

Special cases. Before we proceed to the proof of this statement it will be instructive to look at the simple cases. What is Q_1? O_1 is a line segment bounded by A_1 and B_1, of which A_1 is coloured black and B_1 white; there are no σ_{n-2}'s. To get Q_1 we match A_1 with a point interior to the segment, and we have as a result the configuration already described of a circle with a line segment attached to it at one end.

For Q_2, we see that O_2 is a square bounded by the quadrilateral $A_1A_2B_1B_2$; in accordance with our colouring convention we mark as black the sides A_1A_2

and B_1B_2, and the single vertex B_1. To get Q_2 we match A_1A_2 with an interior 1-cell whose boundary coincides with its own; and B_1B_2 with another 1-cell whose end points are B_2 and B_1', where B_1' is distinct from B_1.

Examination of this figure shows that it is spherelike, and that the vertices A_1, A_2, B_1 and B_2 are all regular, having complements which are Q_1's; but that the lines A_1A_2 etc. are irregular. Since these lines contain no vertices of Q_2, it follows that Q_2 is O-semiregular.

Proof that Q_n is $(n-2)$-semiregular. Lemma 2 states that we can make the proof by showing that the complement of each vertex is regular or $(n-3)$-semiregular, and spherelike; and the discussion of Q_2 brought out the fact that the complement of each doubtful vertex was a Q_1. This suggests at once the method of proof, namely to show that in Q_n the complement of each boundary vertex is a Q_{n-1}, and to show independently that each Q_p is spherelike; then by induction, since we have seen Q_2 to be O-semiregular, we can deduce Q_n to be $(n-2)$-semiregular.

We need to make a slight sharpening of Lemma 2: it is only necessary to prove that each vertex has a spherelike complement and that each vertex except one has an $(n-3)$-semiregular complement. For the latter is enough to show that every σ_p for $n-1 > p > 0$ is regular; any σ_p in this range must have at least one vertex which is not the exceptional one, and consideration of the complement of some other vertex will prove its regularity.

Now in Q_n the vertices not on the boundary, nor matched with the boundary, are clearly regular in all respects, and we need only show

(a) that every boundary vertex except one has an $(n-3)$-semiregular complement,

(b) that every boundary vertex has a spherelike complement.

THEOREM 1. *Every boundary vertex of Q_n, except possibly B_1, has a complement which is a Q_{n-1}.*

No vertex other than B_1 is matched with an interior vertex. Let us consider C_q; we can take its complement first in O_n: this is an O_{n-1}: the complement in Q_n, provided $C_q \neq B_1$, can be got by matching certain boundary simplexes of this O_{n-1} with interior simplexes; in fact in the complement σ_p will be matched to an interior simplex if and only if $C_q\sigma_p$ was matched to an interior simplex in Q_n. Now if $C_q = A_q$ the demonstration is simple: σ_{n-2} is black in O_{n-1} if $A_q\sigma_{n-2}$ is black in O_n; that is if σ_{n-2} contains an even number of B's. And σ_{n-3} is black in O_{n-1} if the number of B's previous to the missing vertex in $A_q\sigma_{n-3}$ is odd; this rule will be the same in O_{n-1},—the q^{th} vertex not, of course, being counted as missing. In fact, in the O_{n-1}, to get the complement of A_q we simply make the colourings or matchings prescribed for a Q_{n-1}.

If C_q is B_q, $q > 1$, in order to bring out clearly that the complement is a Q_{n-1}, rename A_{q-1} and B_{q-1} in the O_{n-1} as B'_{q-1} and A'_{q-1} respectively. This means that the number of B's in a sequence up to a number $\geqq q-1$ is changed from odd to even, or even to odd. So that if $B_q\sigma_{n-2}$ contains an even number

of B's, so does σ'_{n-2}; and if $B_q\sigma_{n-3}$ has an odd number previous to the missing vertex, σ'_{n-3} also has; and the complement of B_q, $q > 1$, is again a Q_{n-1}. This proves Theorem 1.

THEOREM 2. *Q_n has the homology characters of an n-sphere.*

The essential facts about Q_n that we use in the proof are:—first, that it is an O_n with certain boundary simplexes matched with interior simplexes (coloured black); secondly that each black σ_{n-1} has at least one white face—the face got by omitting C_1; and thirdly that one and only one black σ_{n-1} has no black face,—namely $A_1A_2 \cdots A_n$.

To find the homology characters of Q_n we observe first that those of O_n are the characters of a point, $R_p = \delta_0^p$, $\theta_p = 0$; then we consider which chains of O_n which are not cycles become cycles in Q_n. Any such chain has as its boundary a chain on the boundary of O_n together with the image of this chain on the interior σ'_{n-1}'s. But for $p \neq n$ this boundary cycle always bounds a chain having the same property; that is, a chain which becomes zero in Q_n: now the original chain is homologous over O_n to this vanishing chain, and consequently the new cycle in Q_n is homologous to zero. So that every Betti number and torsion coefficient for dimensions other than n and O vanishes. For the dimension n we can only get a new cycle in case the boundary of some σ_{n-1} coincides with that of its σ'_{n-1}, and the Betti number will be the number of such pairs. We have already said that in Q_n there is one such pair. As a result we know that Q_n has the characters of an n-sphere.

From Theorems 1 and 2 we have all that we need for the induction except that we require one further fact,—that the complement of B_1 is spherelike. This consists of two portions: first the complement of B_1 in O_n, which is an O_{n-1}, with the appropriate matchings; and then the complement of the interior point with which B_1 is matched; this is an $(n-1)$-sphere and it will be joined to the other part along the $(n-2)$-simplex $A_1A_2 \cdots A_{n-1}B_n$. This effect can be reproduced by making, in addition to the indicated colourings of the O_{n-1}, the further colouring of $A_1A_2 \cdots A_{n-1}B_n$ as black and the colouring of all its boundary as white. We then find that the figure possesses all the properties used in the proof of Theorem 2, and we can therefore say that B_1 is itself regular in Q_n.

We now know that if Q_{n-1} is $(n-3)$-semiregular, Q_n is $(n-2)$-semiregular; since Q_2 is O-semiregular, we know that we have constructed an n-dimensional $(n-2)$-semiregular complex, and with it a complete set of examples arising from our two problems.

The independence of intercept uniformity. The last question to be discussed concerns the regularity of a cell complex defined in an abstract fashion. Tucker (loc. cit. 1) states three conditions on such a complex for it to be a manifold:—closure, star, and intercept uniformity.

A cell complex has *closure uniformity* if the boundary of each cell has the homology characters of a sphere of one dimension less than itself. Under this

condition it is possible to define regular subdivision and the first derived complex, which will be simplicial.

Star uniformity involves the star of each cell having the characters of an n-cell. With this condition added we know that all the vertices of the first derived complex are regular.

The complex further has *intercept uniformity* if the intercept between any pair of cells is null-like. The intercept of E_p and E_q is the set (if any) of cells E_r which are faces of E_p and have E_q as a face (or vice versa); it is the intersection of the star of E_q with the closure of E_p. It is null-like if all the cycles bound. If a complex has closure and star uniformity, intercept uniformity imposes the further condition on the first derived that all the 1-simplexes are regular.

This means that we can state a necessary and sufficient condition for the independence of intercept uniformity in terms of the concept developed previously; namely that there exists a simplicial complex which is the first derived of a cell complex and which is zero-semiregular. The solution of the problem depends on a knowledge of O-semiregular complexes, although the final result is obviously best stated as a cell complex which is in its own right an example to prove the independence.

Construction of the example. If a simplicial complex is to be expressible as the subdivision of a cell complex, we must be able to assign to each vertex some cell of the complex and in this way define a one-to-one correspondence between the vertices and the cells. Further, two vertices will be joined by a line if and only if their corresponding cells are incident. Since no two cells of the same dimension are incident, no two vertices which are joined by a line correspond to cells of the same dimension. Also, since the boundary of each cell must be a cycle having the characters of a sphere, if a vertex is to correspond to a p-cell it must be possible to express its complement as the join of two subcomplexes, one of dimension $(p - 1)$ and one of dimension $(n - p_{-1})$. These conditions on the simplicial complex are certainly necessary ones.

The simplest example which has been found to satisfy them may be described as follows:—Take a two-sphere and mark on it two lines AB and BC, having one end point in common; match BA with BC. This gives a "pinched sphere": it has the characters of a two-sphere but is not regular. Now consider the join of this pinched sphere with a 1-sphere, or circle, S_1. The irregular set of the pinched sphere is the line AB and both its end points; the closure of the irregular set of the join is the join of AB to S_1, but on this S_1 itself is a regular subset. That is, the closure of the irregular set is a 3-cell; its regular subset is the 1-sphere S_1, lying on its boundary.

In a simplicial subdivision of this whole point set, which is to be O-semiregular, the irregular set must be described as the sum of simplexes, all of whose vertices lie on S_1. This is by no means difficult: we can cover the boundary of the 3-cell in almost any way by 2-simplexes and find a simple cycle of 1-simplexes containing all the vertices of this 2-complex; then by performing the

subdivision so that this cycle coincides with S_1, and by using these vertices to extend the subdivision in any arbitrary way to the 3-cell, and from that outward over the whole set, we will arrive at a O-semiregular complex. But it will not in general be possible to express this complex as the first derived of a cell complex.

First, no vertex lying on S_1 can correspond to a 2-cell of the complex; for its complement can not be expressed as the join of two 1-dimensional sets. Then the vertices of S_1 must correspond to 0-cells, 1-cells, 3-cells and 4-cells of the cell complex. If A_1 corresponds to a 1-cell and lies on S_1, since its complement will be the join of two points on S_1 to a pinched sphere, those two points will have to correspond to vertices of the cell complex, call them A_0^1 and A_0^2. By a dual argument, if A_3 corresponds to a 3-cell and is on S_1, its neighbours on S_1 will be A_4^1 and A_4^2, corresponding to 4-cells. This argument determines a minimum possible subdivision of the 3-cell to satisfy the conditions. We may express it in this way:—the 3-cell is a 3-simplex $A_0^1A_0^2A_4^1A_4^2$ with $A_0^1A_0^2$ subdivided at A_1 and $A_4^1A_4^2$ at A_3; by this the 3-cell is the sum of four 3-simplexes.

Now the rest of the point set is completely regular and clearly we can extend this subdivision to cover the whole set in such a way that the final complex can in its entirety be expressed as the first derived of a cell complex. In order to present a final argument, we give a cell complex constructed to be the one whose first derived is such a complex as has just been described; whether or not this is the case is hard to test and immaterial, as elementary calculation verifies that it has closure and star uniformity but not intercept uniformity. For instance, the intercept of E_1^1 and E_3^1 is not null-like; this corresponds to the fact that, in the simplicial complex described above, the 1-simplex A_1A_3 was irregular.

The required cell complex.

$$
\begin{aligned}
F(E_4^1) &= E_3^1 + E_3^2 - E_3^3\\
F(E_4^2) &= E_3^1 + E_3^4 - E_3^5\\
F(E_4^3) &= E_3^2 - E_3^3 - E_3^4 + E_3^5
\end{aligned}
$$

$$
\begin{aligned}
F(E_3^1) &= E_2^1 - E_2^2 - E_2^3 + E_2^4\\
F(E_3^2) &= -E_2^1 + E_2^2 + E_2^5\\
F(E_3^3) &= -E_2^3 + E_2^4 + E_2^5\\
F(E_3^4) &= -E_2^1 + E_2^3 + E_2^6\\
F(E_3^5) &= -E_2^2 + E_2^4 + E_2^6
\end{aligned}
$$

$$
\begin{aligned}
F(E_2^1) &= -E_1^1 + E_1^2 + E_1^3\\
F(E_2^2) &= -E_1^1 + E_1^2 + E_1^4
\end{aligned}
$$

$$F(E_2^3) = -E_1^1 + E_1^3 + E_1^5$$
$$F(E_2^4) = -E_1^1 + E_1^4 + E_1^5$$
$$F(E_2^5) = E_1^3 - E_1^4$$
$$F(E_2^6) = E_1^2 - E_1^5$$

$$F(E_1^1) = E_0^1 - E_0^2$$
$$F(E_1^2) = -E_0^2 + E_0^3$$
$$F(E_1^3) = E_0^1 - E_0^3$$
$$F(E_1^4) = E_0^1 - E_0^3$$
$$F(E_1^5) = -E_0^2 + E_0^3$$

PRINCETON UNIVERSITY.